

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:10:22 ; Search time 45.5073 Seconds  
(without alignments)  
4345.419 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCLQSTRVRLRENGQHPFST.....MIVKRVSTRSASSQSSR 2057

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10628	99.0	2058	2 A59267	myosin X - human
2	10011	93.3	2052	2 A59297	myosin X - mouse
3	9861	92.8	2052	2 T18519	myosin X - bovine
4	1921	17.9	2357	2 A59249	class VII unconven
5	1840.5	17.2	2215	2 T30870	myosin VIIa - mous
6	1833	17.1	2175	2 A59255	myosin VIIa, long
7	1716	16.0	2098	2 T25888	hypothetical prote
8	1702.5	15.9	1203	2 A59257	myosin VIIa, short
9	1687	15.8	2121	2 A59233	myosin VII-like pr
10	1593	14.8	3511	2 A59295	unconventional myo
11	1589.5	14.8	3530	2 A59266	unconventional myo
12	1512.5	14.1	1529	2 A59310	unconventional myo
13	1504.5	14.0	1846	2 A59289	myr 6, unconventi
14	1500	14.0	2116	2 A26655	myosin heavy chain
15	1492.5	13.9	2022	2 A59256	myosin-IXD ismilia
16	1492	13.9	2548	2 E59435	myosin IXA (import
17	1490.5	13.9	2245	2 T18278	myosin heavy chain
18	1487.5	13.9	1611	2 A84743	probable myosin he
19	1487	13.9	2626	2 T31099	myosin-Rhocap prot
20	1466	13.8	1783	2 T42386	unconventional myo
21	1465	13.8	1260	2 T14276	myosin-like protei
22	1463	13.8	1528	2 T14276	myosin-like protei
23	1471	13.7	1980	2 S54307	myosin heavy chain
24	1461	13.6	1855	2 A59254	myosin heavy chain
25	1458.5	13.6	1477	2 T00957	myosin heavy chain
26	1458	13.6	1520	2 S46444	myosin MYA1, class
27	1458	13.6	1583	2 T00727	myosin heavy chain
28	1457.5	13.6	1628	2 B59254	myosin heavy chain
29	1450.5	13.5	1556	2 F96857	hypothetical prote

30	1442	13.4	1574	1 A38454	myosin MYO2 - yeast
31	1435	13.4	1515	2 S51824	myosin heavy chain
32	1434.5	13.4	1446	2 T04528	myosin heavy chain
33	1434	13.4	1853	1 A46761	myosin heavy chain
34	1430.5	13.3	1490	2 B4726	probable unconvent
35	1429.5	13.3	1830	1 S19188	myosin-V - chicken
36	1424.5	13.3	1242	2 T45976	myosin heavy chain
37	1423	13.3	1374	2 D85390	myosin-like protei
38	1421.5	13.2	1375	2 T05200	myosin heavy chain
39	1399.5	13.0	1502	2 T14278	myosin-like protei
40	1399.5	13.0	1736	2 F86178	hypothetical prote
41	1399	13.0	1516	2 T41235	probable myosin he
42	1394	13.0	1502	2 D84587	probable myosin he
43	1379	12.9	1471	2 T40117	myosin-2 isoform -
44	1373.5	12.8	1509	1 A27224	myosin heavy chain
45	1370.5	12.8	1792	2 T13939	myosin V - fruit f

#### ALIGNMENTS

##### RESULT 1

A59267  
myosin X - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Sep-2000  
C:Accession: A59267  
R:Rogers, M.S.; Streibler, B.E.  
Submitted to GenBank, March 2000  
A:Description: Identification of myosin X as a specific binding partner for the tumor se  
A:Reference number: A59267  
A:Accession: A59267  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2058 <ROG>  
A:Cross-References: GB:AF234532; P1DN:AA37875.1  
A:Genetics:  
A:Gene: MYO10  
C:Superfamily: myosin motor domain homology  
F:66-727/Domain: myosin motor domain homology <MWO>

Query Match	99.0%	Score 10628	DB 2	Length 2058	Best Local Similarity 99.3%	Pred. No. 0	Matches 2043	Conservative 5	Mismatches 7	Indels 2	Gaps 2
QY	1	FCLQSTRVRLRENGQHPFSTVNSCAEGIVRRTYGVPTTKOSTHOKVTAMHPTNEE	60								
DB	4	FTEBSTRVRLRENGQHPFSTVNSCAEGIVRRTYGVPTTKOSTHOKVTAMHPTNEE	63								
QY	61	GVDDWASLTSLHGGSIMYNLFORYKRNQIMTYIGSILASVNPYQPIAGLYEPATWEOYSR	120								
DB	64	GVDDWASLTSLHGGSIMYNLFORYKRNQIMTYIGSILASVNPYQPIAGLYEPATWEOYSR	123								
QY	121	RHLGELPHRIFAIANECYRCILMKRHDNOCILIKESGAGKTESTYLILKPLSVISQGLE	180								
DB	124	RHLGELPHRIFAIANECYRCILMKRHDNOCILIKESGAGKTESTYLILKPLSVISQGLE	183								
QY	191	LSLKKTSQVRRATLIESPIMEARGNATVYNNNSRRGKTVQNLICCKGNIGGRITDC	240								
DB	194	LSLKKTSQVRRATLIESPIMEARGNATVYNNNSRRGKTVQNLICCKGNIGGRITDC	243								
QY	241	ILSSQNRVVRONPGBRNHIFVALLAGLEHREBEFYSTBENHYHNLQSGCVEDKTISD	300								
DB	244	LL-ENKRVVRONPGBRNHIFVALLAGLEHREBEFYSTBENHYHNLQSGCVEDKTISD	302								
QY	301	QESFEVITAMDVNQFSKEVREVSRLIAGILHAGNEFTTAGAQVSPKTAGSABLL	360								
DB	303	QESFEVITAMDVNQFSKEVREVSRLIAGILHAGNEFTTAGAQVSPKTAGSABLL	362								
QY	361	GLDPQTLDALTGSMPLRGSEILTPLVNQAVDSRSLAALVACCEBWTIKKINSRIK	420								
DB	363	GLDPQTLDALTGSMPLRGSEILTPLVNQAVDSRSLAALVACCEBWTIKKINSRIK	422								

QY	421	GNEDFSGICILDI	FGENENVNH	PEOFNINYNNEKLOE	FNGHITS	LEBLOEX	SRGGLWE	480	
Db	422	GNEDFSGICILDI	FGENENVNH	PEOFNINYNNEKLOE	FNGHITS	LEBLOEX	SRGGLWE	482	
QY	481	DIWDIWDGECJDL	IEKJGLILALINEESH	PPAOTDSTLLEKLSOHANH	FVYKERVAVN			540	
Db	483	DIWDIWDGECJDL	IEKJGLILALINEESH	PPAOTDSTLLEKLSOHANH	FVYKERVAVN			542	
QY	541	NGVGVYAGEVOY	DVARGILEKORDTF	RODDLNLABSRPDIYDL	FEHVSSNNODTLKC			600	
Db	543	NGVGVYAGEVOY	DVARGILEKORDTF	RODDLNLABSRPDIYDL	FEHVSSNNODTLKC			602	
QY	601	GSKHRPTVSSOPK	VDLSHLSMAATLSSNSNPFVRCI	KPMQOMQPOFOA	VVLNLOLRYSG			660	
Db	603	GSKHRPTVSSOPK	VDLSHLSMAATLSSNSNPFVRCI	KPMQOMQPOFOA	VVLNLOLRYSG			661	
QY	661	MLETVIRIKAGAVAR	PFQDFYKRYKYLMRNLT	PEDVAGKCTSLQI	YDASNSEMQJGK			720	
Db	662	MLETVIRIKAGAVAR	PFQDFYKRYKYLMRNLT	PEDVAGKCTSLQI	YDASNSEMQJGK			721	
QY	721	TKVPLRESLEQLEK	RREBEVSHAAMVIRAHVGLF	LARKQYRKVLYCVIIT	IOKNYRAPLL			780	
Db	722	TKVPLRESLEQLEK	RREBEVSHAAMVIRAHVGLF	LARKQYRKVLYCVIIT	IOKNYRAPLL			781	
QY	781	RRRFLHKKAAIV	FPQKOLRGQIARVYVQLLA	EXKQOEKKQOEBEKKQREERERER				840	
Db	782	RRRFLHKKAAIV	FPQKOLRGQIARVYVQLLA	EXKQOEKKQOEBEKKQREERERER				841	
QY	841	ERRRABELRAQOE	EEFRKQOELEALOKSQEAL	ITRELEKQENKQVEITLREK	IEDIQ			900	
Db	842	ERRRABELRAQOE	EEFRKQOELEALOKSQEAL	ITRELEKQENKQVEITLREK	IEDIQ			901	
QY	901	RNKEQOELESLT	ASLOKQERRDOELR	LEBACBAQOEFLS	LNPFIDEICVNR	IERSL		960	
Db	902	RNKEQOELESLT	ASLOKQERRDOELR	LEBACBAQOEFLS	LNPFIDEICVNR	IERSL		961	
QY	961	SGGSFSSSETLAS	ACEEKNFNFSPQYPREE	DEBGFADDDAFKOSP	PNPSEHGSHDQRTS			1020	
Db	962	SGGSFSSSETLAS	ACEEKNFNFSPQYPREE	DEBGFADDDAFKOSP	PNPSEHGSHDQRTS			1021	
QY	1021	GIRTSDDSEEDP	PNYNDTVPTSPASDT	TVLAPSVQDSGLHNS	SGESTYCMQONAGD			1080	
Db	1022	GIRTSDDSEEDP	PNYNDTVPTSPASDT	TVLAPSVQDSGLHNS	SGESTYCMQONAGD			1081	
QY	1081	LPSPDGDYDYDD	YEDGAI	TSGSVTPBNSYGSQMS	PDYRCSVGTYNSSGAYRESGA			1140	
Db	1082	LPSPDGDYDYDD	YEDGAI	TSGSVTPBNSYGSQMS	PDYRCSVGTYNSSGAYRESGA			1141	
QY	1141	OSSFDDSEEDP	FSRPTDDEL	SYRDSVYCTTLY	FHSFLYMKGLN	SMWRKRCVYKD		1200	
Db	1142	OSSFDDSEEDP	FSRPTDDEL	SYRDSVYCTTLY	FHSFLYMKGLN	SMWRKRCVYKD		1201	
QY	1201	ETFLWFRSKQEL	KQGM	LHKKGSGSTLS	RWMKRWFLRQSK	LYMPENDEEKLKGTV		1260	
Db	1202	ETFLWFRSKQEL	KQGM	LHKKGSGSTLS	RWMKRWFLRQSK	LYMPENDEEKLKGTV		1261	
QY	1261	EVRTAKEI	IIDNTTKENG	IDIIMADDTF	PHLIESP	DASQWFSVL	SQVHASTDOEIOEMHD	1320	
Db	1262	EVRTAKEI	IIDNTTKENG	IDIIMADDTF	PHLIESP	DASQWFSVL	SQVHASTDOEIOEMHD	1321	
QY	1321	EOANPQNAVGL	TDVGLIDS	VCASDSDPR	PNSFVITITAN	VILHCNAD	TEPEMHMTITLLQR	1380	
Db	1322	EOANPQNAVGL	TDVGLIDS	VCASDSDPR	PNSFVITITAN	VILHCNAD	TEPEMHMTITLLQR	1381	
QY	1381	SKGDFRVGQET	IVGMLHKEVNS	PKWSSLK	KRWFL	LTNSLDYK	SSSRKNA	1440	
Db	1382	SKGDFRVGQET	IVGMLHKEVNS	PKWSSLK	KRWFL	LTNSLDYK	SSSRKNA	1441	
QY	1441	LVANSLC	SVVPDEK	IKFETGYMNT	VYGRKICYRL	YTLKL	NEATRMS	SVIONVDTKAP	1500
Db	1442	LVANSLC	SVVPDEK	IKFETGYMNT	VYGRKICYRL	YTLKL	NEATRMS	SVIONVDTKAP	1501
QY	1501	IDTPTQOOL	IODIKENCLNS	DVEQIYK	KNPILYRTHHP	LHSEPLR	PYGDIN	MLNLKDG	1560

Db 1502 IDTTLQDLOIKENCANSDVBEIYKRNPIKRTTHPLPSPLLPLPYGININLNLKDKG 1561

Qy 1561 YTTLQDAIKIENSLOLESMSDPIPIIGILQTHDLRLPLRDELYCOLIKQTNKVPHG 1620

Db 1562 YTTLQDAIKIFNSLOLESMSDPIPIIGILQTHDLRLPLRDELYCOLIKQTNKVPHG 1621

Qy 1621 SVGNLYSMQOLLITCLSCFPLBSRGLIKYKFKHRIKRPQPGSTMEKALTYSIKTKC 1680

Db 1622 SVGNLYSMQOLLITCLSCFPLBSRGLIKYKFKHRIKRPQPGSTMEKALTYSIKTKC 1681

Qy 1681 REFPPSRDEIALIHRQEMTSTVYCHGSGCKITINSHTAGVEVEKLIGLAMEDSRM 1740

Db 1682 REFPPSRDEIALIHRQEMTSTVYCHGSGCKITINSHTAGVEVEKLIGLAMEDSRM 1741

Qy 1741 FALFEYNGHVDKAIESTTVADVLAKEKLAISEVBDLPWKYFKLYCTFLDNDNPKDS 1800

Db 1742 FALFEYNGHVDKAIESTTVADVLAKEKLAISEVBDLPWKYFKLYCTFLDNDNPKDS 1801

Qy 1801 VEPFPMFEQAEVATIHGHHPAPRENTQVLAALROYLOSGDYTLHAAPLPEEYSLQRLK 1860

Db 1802 VEPFPMFEQAEVATIHGHHPAPRENTQVLAALROYLOSGDYTLHAAPLPEEYSLQRLK 1861

Qy 1861 ARISQSTKFTPCERLKERKRTSPLEGTLRRSPFRGSVVRQVBEQMLDMWIKBEVSSAR 1920

Db 1920 ARISQSTKFTPCERLKERKRTSPLEGTLRRSPFRGSVVRQVBEQMLDMWIKBEVSSAR 1921

Qy 1921 ASITDKRRKQGNNOEQAMAKYMALIKEMPGYSTLFDFVECKSGGFPQELMLGVSDAYS 1980

Db 1922 ASITDKRRKQGNNOEQAMAKYMALIKEMPGYSTLFDFVECKSGGFPQELMLGVSDAYS 1981

Qy 1981 VYKRGGRPLPEVFOYEHIILSFQAPLANTYKIVDERELLETSEVNDVALKRAYTSMIV 2040

Db 1982 VYKRGGRPLPEVFOYEHIILSFQAPLANTYKIVDERELLETSEVNDVALKRAYTSMIV 2041

Qy 2041 KKRYSYTRRSASQSSSR 2057

Db 2042 KKRYSYTRRSASQSSSR 2058

RESULT 2

A59297

myosin X - mouse

C:Species: Mus musculus (house mouse)

C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000

C:Accession: A59297

R:Yonezawa, S.; Kimura, A.; Koshida, S.; Maseaki, S.; Ono, T.; Hanai, A.; Sonta, S.; Kagee Biochem. Biophys. Res. Commun. 271, 526-533, 2000

A:Title: Mouse myosin X: Molecular architecture and tissue expression as revealed by nor A:Reference number: A59297

A:Accession: A59297

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2062 <YON>

A:Cross-references: GB:AJ249706; PIDN:CAB56466.2

A:Experimental source: strain BALB/c; tissue 1lb testis

C:Genetics:

A:Gene: myo10

C:Superfamily: myosin motor domain homology

F:66-727/Domain: myosin motor domain homology <MMO>

Query Match 93.3%; Score 10011; DB 2; Length 2062;

Best Local Similarity 93.3%; Pred. No. 0;

Matches 1925; Conservative 64; Mismatches 64; Indels 10; Gaps 6;

Qy 1 FCLQSTRVWLRENGQHPFSTVNSCAEGIVFRTDYGVFTYKOSTITTHQKVTAMHPPTNEE 60

Db 4 FPEFGARVWLRENGQHPFSTVNSCAEGIVVFQTDYGVFTYKOSTITTHQKVTAMHPPTNEE 63

Qy 61 GVDMDASLTLEHGSIMVYNLFQYKRNQITVYIGSLIASVNPQPLAGLVEPATMEOYSR 120

Db 64 GVDMDASLTLEHGSIMVYNLFQYKRNQITVYIGSLIASVNPQPLAGLVEPATMEOYSR 123



Qy	121	RHIGELPHRIPALINECEYRCMLKXNDHNCILIKESGAGKTESRKLILKPLSVISQOOLE	180
Db	124	CHIELEPPIHAIANECEYRCMLKXNDHNCVILISBGSAGKTESRKLILKPLSVISQOOLD	183
Qy	181	LSLKEKTSVCYRALLESSEPIMEAEAGNATVYVNNSSRPGKVFVQNI'COKGI'OGGRIVDC	240
Db	184	LGLOEKTSVSQALIOSSPIMEAEAGNATVYVNNSSRPGKVFVQNI'COQGNIOGGRIVDY	243
Qy	241	ILSSONRVVRONPGEERNHIFVYALLAGLHEHEENEFPYLSTEBNHYVINOQSGVEDKTIISD	300
Db	244	LL-EKNRVVRIRIPGERNHYIFVYALLAGLHDQEREHEEFPYLSTEBNHYVINOQSGVEDKTIISD	302
Qy	301	QESREPVITANDWQFSKEBEVRSRLIAGLILHIGNEFTIAGAGNAQVSPKALCRSEJEL	360
Db	303	QESREPVITANEVWQFSKEBEVRLIAGLILHIGNEFTIAGAGNAQVSPKALCRSADLL	362
Qy	361	GLDPTOLDALTORSMPFRGEEILTPLYVQOAV'OSRDSIALAALVACCEPWYIKKINSRIK	420
Db	363	GLDPTOLDALTATONSMLRGEELITPLSVQOAVDSRDSIALAALVACCEPWYIKKINSRIK	422
Qy	421	GNEDFKSIGLIDIFGFENFEVNHPEQFINYANAEKLOEYFNKHI'PSLEOLEYSREGLWE	480
Db	423	GKDPFKSIGLIDIFGFENFEVNHPEQFINYANAEKLOEYFNKHI'PSLEOLEYSREGLWE	482
Qy	481	DIDWINDNECDLLEKKLGLALLINESHPFOADTSTLEKLSGHANHHYVYKPRAVN	540
Db	483	DIDWINDNECDLLEKKLRLALINEESHFOADTSTLEKLSGHANHHYVYKPRAVN	542
Qy	541	NFGVGHVAGEVOYDVRGIL'ENKRDTPFRDNLNLT'RESRFDTIYDL'FEHVSRNNODTLKC	600
Db	543	NFGVGHVAGEVOYDVRGIL'ENKRDTPFRDNLNLT'RESRFDTIYDL'FEHVSRNNODTLKC	602
Qy	601	GSKRRP'VSSQFKVDSLHSLMATLSSNPFVVC'CIKKNMO'KMPDQFOA'VVLNQLNLYSG	660
Db	603	GSKRRP'VSSQFK-DSLHSLMATLSSNPFVVC'CIKKNTO'KMPDQFOQVVLNQLNLYSG	661
Qy	661	MLETYRIRIKAGYAVRRPFDYKRYKXVLMRNIAL'PEDVRGKSTSLLOLYDANSBEMOJLK	720
Db	662	MLETYRIRIKAGYAVRRPFDYKRYKXVLMRNIAL'PDIDRGKCTVLLQVYDANSBEMOJLK	721
Qy	721	TKVFLRELEBOKLERREBEVSHAMV'RAVUL'FLAKQORYKLYLCVILIOKNYRPAFL	780
Db	722	TKVFLRELEBOKLERREBEIDRAMV'RAHIL'GLAKQORYKLYLCGVITIOKNYRPAFL	781
Qy	781	RRRFLHLKKAIVFOQOLRGOI'ARRVYRQOLAE'KEOEKKEKKEOEKKEKKEEEREER	840
Db	782	RKKFLHLKKAIVPOKOLRGOLARVYRQOLAE'KELEKKEKKEEER--KKEEEREEREER	839
Qy	841	ERRAE-IRAOOEBETRKQOELI'EAOLKSQKEAE'LTRELEKOKENKQVEILTRLEKEIDL	899
Db	840	AQREADLLRAHQEATRRQOELI'EAOLKSQRAADLTRELEKORENKQVEILTRLEKEIDL	899
Qy	900	QRMKEQOELISTEASLOQLERRODELARLEEA'CRAOERLESINPEIDECRNIERS	959
Db	900	QRMKEQOELISTEALQOLRBEILRLLEDEACRAOERLESINPEIDECRNIERS	959
Qy	960	LSGSEFS-----SLIASASACEKPFNFNSOPV--BEVDEGFEADDDAFKOSPMPSEHG	1014
Db	960	LSVSGEISIGEBELSELASASGEKSPFNFSOPVPEEVEDEGFEADDDAFKOSPMPSEHG	1019
Qy	1015	SDORTSGIRTSDDSSSEBDPYMNDIVPTSPSADSVLLAPSVODSGSLHNSSGSESTYCM	1074
Db	1020	SDORTSGIRTSDDSSSEBDPYMNDIVPTSPSADSVLLAASGQDASLHNSSGSESTYCM	1079
Qy	1075	PONAGDLSPDGDYVYDDDYEDGATITGSGSVTF'FNSN'GSGQWSPYRCSVGTNNSGAYR	1134
Db	1080	PONPBDLSPDGDYVYDDDYEDGATITGSGSVTF'FNSN'GSGQWSPYRYSVGTNNSGAYR	1139
Qy	1135	FSSEGAQSSFEDESEBDFOSRFDTDDEL'SYRDSV'SCTLLPYFHSFLYMKGLMNSMGR	1194
Db	1140	FSSBEAQSSFEDESEBDFOSRFDTDDEL'SYRDSV'SCTLLPYFHSFLYMKGLMNSMGR	1199
Qy	1195	WCVLNDETFWLRSHQOALKQGMHKKGGSGSTLR'RWKQRMFVLRQSKLMTYENDSEE	1254

Db 1200 WCVLKDITFLMFPSKQBALQGMHLKKGSSSTLSRNMKKRPFVLRQSLMTFFENDSEE 1259

Qy 1255 KLKGTVEVTRAKELIDNTTKENGIDIIIMADRTFHLIAESPEDASQWFSVLQVHASTDQE 1314

Db 1260 KLKGTVEVTRAKELIDNTSKENGIDIIADRTFHLIAESPEDASQWFSVLQVHASTDQE 1319

Qy 1315 IQEMHDAQNPONAAGTLDVGLIDSVCASSSPRPNSFVITANRVLHCADPTPEMHHV 1374

Db 1320 IREHMDQANPONAAGTLDVGLIDSVCASSSPRPNSFVITANRVLHCAYTPEEMHHV 1379

Qy 1375 IYLLQSKGTTRVEGGEFIVRGMLHKEVKNSPKMSSIKLKKRPFVLTNNSLDYKSEKN 1434

Db 1380 IYLLQSKGTTRVEGGEFIVRGMLHKEVKNSPKMSSIKLKKRPFVLTNNSLDYKSEKN 1439

Qy 1435 ALKLGTVLNLSCSVPPDEKIFKETGVNVTYYGRNGCTRLTYKTLNEATRNSSVIONV 1494

Db 1440 ALKLGTVLNLSCSVPPDEKIFKETGVNVTYYGRNGCTRLTYKTLNEATRNSSVIONV 1499

Qy 1495 TDYKAPIDPTQOQLQDIKENCLNSDVVEOITYRNPILRTYTHHPHLSPLLPLPYGGINLN 1554

Db 1500 TDYKAPIDPTQOQLQDIKENCLNSDVVEOITYRNPILRTYTHHPHLSPLLPLPYGGINLN 1559

Qy 1555 LLYKKGTTTLODEKIKTFNSLQQLESMSDPIPIIOGLQGHLPRLRDELVYQOLKQTN 1614

Db 1560 LLYKKGTTTLODEKIKTFNSLQQLESMSDPIPIIOGLQGHLPRLRDELVYQOLKQTN 1619

Qy 1615 KVPHPGSGVNLVSMQIITLCSCTFLPSRGILKYLKFKLAKIREQPPTEMEKVALFTYES 1674

Db 1620 KVPHPGSGVNLVSMQIITLCSCTFLPSRGILKYLKFKLAKIREQPPTEMEKVALFTYES 1679

Qy 1675 LKKTCKREFVPSRDEIBALIHROEMTSVYCHGGGSCKITINSHTTAGEVVEKILRGLAM 1734

Db 1680 LKKTCKREFVPSRDEIBALIHROEMTSVYCHGGGSCKITINSHTTAGEVVEKILRGLAM 1739

Qy 1735 EDSNNMMLPEYNGHVVKALIESRTVADVLAKEBKLAATSFVGDLPMKFFPKLYCFLDITD 1794

Db 1740 EDSNNMMLPEYNGQVVKALIESRTVADVLAKEBKLAATSFVGDLPMKFFPKLYCFLDITD 1799

Qy 1795 NVPPDSVEFPMFPEQAHAEVYHGHHPAPENLQVLAALRLQYLOGDYTLAAIPLPEEVY 1854

Db 1800 SMPDQGVFPMFPEQAHAEVYHGHHPAPESLOYLAALRLQYLOGDYTLPHTSIPPLEEVY 1859

Qy 1855 SLQRLKARISOSTKTFPTPCERLEKRTSPLEGLTRRSFRTGSVVRQKVEEQLMDWIKI 1914

Db 1860 SVQRLKARISOSTKTFPTPERLEKRTSPLEGLTRRSFRTGSVVRQKVEEQLMDWIKI 1919

Qy 1915 EVSSARASIIDKRRKFGANQOQAMAKYMLIKEMPGYSTLPDYCKEKGFPQELMLGV 1974

Db 1920 EVCSARASIIDKMKKQGMTOEQAMAKYMLIKEMPGYSTLPDYCKEKGFPQELMLGV 1979

Qy 1975 SADAVSYKKGEGRPLEVFOYEHLSFGAPLANTYKLVDERELLFETSEVNVVAKLMKA 2034

Db 1980 SAEVAVSYKKGEGRPLEVFOYEHLSFGAPLANTYKLVDERELLFETSEVNVVAKLMKA 2039

Qy 2035 YISMIYKKRYSTTRSSAGSSSR 2057

Db 2040 YISMIYKKRYSTTRSSAGSSSR 2062

RESULT 3

T18519

myosin X - bovine

C.Species: Bos primigenius taurus (cattle)

C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Feb-2001

C.Accession: T18519

R.Corey, D.P.; DeZilker, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.

submitted to the EMBL Data Library, April 1996

A.Description: Cloning and expression of myosin X, a novel unconventional myosin with p

A.Reference number: Z18942

A.Accession: T18519

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A:Residues: 1-2052 <COR>  
A:Cross-references: EMBL:U55042; NID:G1755048; PID:G1755049; PIDN:AAB39486.1  
A:Experimental source: aorta  
C:Superfamily: myosin motor domain homology  
C:Keywords: nucleotide binding; P-loop  
F:66-727/Domain: myosin motor domain homology <MOC>  
F:157-164/Region: nucleotide-binding motif A (P-loop)

Query Match 92.8%; Score 9661; DB 2; Length 2052;  
Best Local Similarity 92.7%; Pred. No. 0;  
Matches 1909; Conservative 76; Mismatches 62; Indels 12; Gaps 6;

1 FCLQGRVWLRNGQHFPSTVNSCAGIIVFRTDYQVFTYKOSTITHQKVTAMAPTNEE 60  
4 FPEGRVWLRNGQHFPSTVNSCAGIIVFRTDYQVFTYKOSTITHQKVTAMAPTNEE 63  
61 GVDMASTLTHNGGSLMNLVFORKRNQITWYIGSLASVNPQIPAGLYEPTMQR 120  
64 GVDMASTLTHNGGSLMNLVFORKRNQITWYIGSLASVNPQIPAGLYEPTMQR 123  
121 RHUGELPHIIPAIANECYRCIMGRHNOCLIKGESGAGKTESTKILKPLSVISQOSLE 180  
124 CHGELPPIHFAIANECYRCIMGRHNOCLIKGESGAGKTESTKILKPLSVISQOSVD 183  
181 LSIKERTSCVERALLESSPIMEAFGNAKTVNNNSRFGKPVQNLIGKGNIGGRIVDY 243  
184 LSKERTSCVERALLESSPIMEAFGNAKTVNNNSRFGKPVQNLIGKGNIGGRIVDY 243  
241 ILSORVVRONGENNYHIFVALLAGLHEBEREYTLSTPREYHNLNOSGCEDTITD 300  
244 LL-EKRRVVRONGENNYHIFVALLAGLHEBEREYTLSTPREYHNLNOSGCEDTITD 302  
301 QESFREVITAMDVQFSKEKEVREVSRLAGILHIGNIIFTTGAQVSEFKALGRASLL 360  
303 QESFREVITAMDVQFSKEKEVREVSRLAGILHIGNIIFTTGAQVSEFKALGRASLL 362  
361 GLDPTQLTALTORSHFLRGEELITPLNVOQAVDSDSLAMALYACCFEMVLIKINSRIK 420  
363 GLDPTQLTALTORSHFLRGEELITPLNVOQAVDSDSLAMALYACCFEMVLIKINSRIK 422  
421 GNEDEFSTIGLDFGFEFENVHFEOPNINYANEKIQEFNKGIFSELBOLEYSEBGLVWE 480  
423 GNEDEFSTIGLDFGFEFENVHFEOPNINYANEKIQEFNKGIFSELBOLEYSEBGLVWE 482  
481 DIDMDINGECLDLIEKGLGLALINEESHFPQATDSTLLEKLSQANNNHFYVKPVAVN 540  
483 DIDMDINGECLDLIEKGLGLALINEESHFPQATDSTLLEKLSQANNNHFYVKPVAVN 542  
541 NFGVKAIVAGEVOYDVAGIIEKNRDPTRDULNLRESRPFYIDLFEEHVSSRNNQDTLKC 600  
543 NFGVKAIVAGEVOYDVAGIIEKNRDPTRDULNLRESRPFYIDLFEEHVSSRNNQDTLKC 602  
601 GSKHRRPTVSSQPKVDSLHSLMATTSSNPFPVRCIKPMQKMPDFDAVVANNQIRYSG 661  
603 GSKHRRPTVSSQPKVDSLHSLMATTSSNPFPVRCIKPMQKMPDFDAVVANNQIRYSG 661  
661 MLETVIRIRAGYAVRRPQDFYKRYKVLNRNLALPEVDGKCTSLQLDASNSEWQLK 720  
662 MLETVIRIRAGYAVRRPQDFYKRYKVLNRNLALPEVDGKCTSLQLDASNSEWQLK 721  
721 TYVFLRESLEQLEKKEEVEVSHAAVITRAHVLGFLAKRQYKRVLYIQNRYAFLL 780  
722 TYVFLRESLEQLEKKEEVEVSHAAVITRAHVLGFLAKRQYKRVLYIQNRYAFLL 781  
781 RRRFLHKKAAIVFOQLAGOTARRVYRQLAEKKEEKKKEEKKKEEKKKEEKKKEEKKKEE 840  
782 RRRFLHKKAAIVFOQLAGOTARRVYRQLAEKKEEKKKEEKKKEEKKKEEKKKEEKKKEE 841  
841 ERRREALRAQEEETKQOELBAL-OKSQKEALITRELEKOKENKOVEEIIILKEIEIDU 899  
842 ERRREALRAQEEETKQOELBAL-OKSQKEALITRELEKOKENKOVEEIIILKEIEIDU 901  
900 QRMKEQOELSTLEASTLQKQERRDDELRLLEBEACRAAQEFLESINPDEIDECVNIERS 959

902 QRMKEQOELSTLEASTLQKQERRDDELRLLEBEACRAAQEFLESINPDEIDECVNIERS 961  
960 LSGSEFSELEAESCEKPNFNFSQAPP-BEEVDEGEADDADAKDSPNPEHSHDOR 1018  
962 LSVGGCTGGE--QGAKEKPFSPNFQPPBEEVEVEGEADDADAKDSPNPEHSHDOR 1019  
1019 TSGIRTSDDSEEDPYMNDTVPTSPADSTYLLAPVQDSGLHNSSGSESTYCPQNA 1078  
1020 TSGIRTSDDSEEDPYMNDTVPTSPADSTYLLAPSE-----HDSAGBPTTCLPQNP 1073  
1079 GDLPSPDGDYDQDDYEDGATSSGSVTFSSNSYSQNSPDYRCGVTYNSGAYRFSSE 1138  
1074 GALPAPBEGDYDQDDYEDGATSSGSVTFSSNSYSQNSPDYRCGVTYNSGAYRFSSE 1133  
1139 GAQSSFEDESEEDPDRPTDDELSTRDSVYSCTLLPYFHSFLYMKGLMNSMKRMCVL 1198  
1134 GAQSSFEDESEEDPDRPTDDELSTRDSVYSCTLLPYFHSFLYMKGLMNSMKRMCVL 1193  
1139 KDEFLMFRSKQBALKQGLHKKGGSSSTLSRNNKKEWFLVRLQSLMYFENDSEBKLG 1258  
1194 KDEFLMFRSKQBALKQGLHKKGGSSSTLSRNNKKEWFLVRLQSLMYFENDSEBKLG 1253  
1259 TVEVTRAKIIDNTYKENGIDITMADRTPHLIAESPEDASQWFSYLSGVHASTDOEIDEM 1318  
1254 TVEVTRAKIIDNTYKENGIDITMADRTPHLIAESPEDASQWFSYLSGVHASTDOEIDEM 1313  
1319 HDEQANPQNAVGTLDVGLIDSVCAQSDSPDRPSPFITANRVLHONACTPEEMHMITLL 1378  
1314 HDEQANPQNAVGTLDVGLIDSVCAQSDSPDRPSPFITANRVLHONACTPEEMHMITLL 1373  
1379 QRSKQDTEVEGEPIVRGMLHKEVNSPMSLSKLKKEWFLVTHNSLDYKSSERXALKL 1438  
1374 QRSKQDTEVEGEPIVRGMLHKEVNSPMSLSKLKKEWFLVTHNSLDYKSSERXALKL 1433  
1439 GTLVNLSLCSVPPPEKIFKETGYANVTYGRKQCYRLYTKLNEATWSSYIQVNTDTK 1498  
1434 GTLVNLSLCSVPPPEKIFKETGYANVTYGRKQCYRLYTKLNEATWSSYIQVNTDTK 1493  
1439 APIDPPTQQLIDIKENCLNSDVVEQIYKRNPLRHTHPLSHPLPPLPYGINLNLKD 1558  
1494 APIDPPTQQLIDIKENCLNSDVVEQIYKRNPLRHTHPLSHPLPPLPYGINLNLKD 1553  
1559 KGYTTLQDEALIKIPNSLQOLESMSDPIPIQIGLQTDGDLRLRLDELVCOLIKQTNKVP 1618  
1554 KGYTTLQDEALIKIPNSLQOLESMSDPIPIQIGLQTDGDLRLRLDELVCOLIKQTNKVP 1613  
1619 PGSVGNLSWQITLCTPLPSRGLKYLKFLHKLRIEOPFGTEMEKALFIYESLTKT 1678  
1614 PGSVGNLSWQITLCTPLPSRGLKYLKFLHKLRIEOPFGTEMEKALFIYESLTKT 1673  
1679 KCRFVPRDEIEALIHQEMTSTYVCHGGGSKTINSHTTAGVVEKTLRGLMEBSR 1738  
1674 KCRFVPRDEIEALIHQEMTSTYVCHGGGSKTINSHTTAGVVEKTLRGLMEBSR 1733  
1739 NMFLAEVNGHVDAKIESRTVADVLAKEFKLAATSEGDLPWKFFYLKCYFLDITDNPV 1798  
1734 NMFLAEVNGHVDAKIESRTVADVLAKEFKLAATSEGDLPWKFFYLKCYFLDITDNPV 1793  
1799 DSVFAPNFEQAHAEAVIHGHHPAPEENIQTALRLQYLOGDYTLHAAIIPLEEYYSIOR 1858  
1794 DSVFAPNFEQAHAEAVIHGHHPAPEENIQTALRLQYLOGDYTLHAAIIPLEEYYSIOR 1853  
1859 LKARISOSTKTPPEERLEKRTSLBEGTLRSRPTGVSVMQKVEEOMLMWIMEEVS 1918  
1854 LKARISOSTKTPPEERLEKRTSLBEGTLRSRPTGVSVMQKVEEOMLMWIMEEVS 1913  
1919 ARAAIIIDMKRFQGNQNOQAAKATYALIKEMPGYSTLFDVECKEGGFPQEBIMLGVSADA 1978  
1914 ARAAIIIDMKRFQGNQNOQAAKATYALIKEMPGYSTLFDVECKEGGFPQEBIMLGVSADA 1973  
1979 VSVYRGGGRPLEVFOYHHIISFGAPLANTYKIVDERELLPETSEVVDAKLMKAYISM 2038

Db 1974 VSVYKKGEGRPLEVFQYEHIIISFCGAPLANTYKIVDERELLFTETSEVVDVAKMKRAYISM 2033

QY 2039 IVKKRYSTTRASASQGSRR 2057  
|||||:|||||  
Db 2034 IVKKRYSTSRVSSQGSRR 2052

## RESULT 4

A:Accession: A59249  
 C:Class: VII unconventional myosin - slime mold (Dictyostellium discoideum)  
 C:Species: Dictyostellium discoideum  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Sep-2000  
 C:Accession: A59249  
 R:Titus, M.A.; Kuspa, A.; Loomis, W.F.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994  
 A:Title: Discovery of myosin genes by physical mapping in Dictyostellium.  
 A:Reference number: 220873; PMID:95023928; PMID:7937787  
 A:Accession: A59249  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-2357 <TTT>  
 A:Cross-references: GB:J35321; NID:96226760; PIDN:AAF06035.1; PID:96226761  
 A:Experimental source: gene myoI; product class VII unconventional myosin  
 R:Titus, M.A.  
 Curr. Biol. 9, 1297-1303, 1999  
 A:Title: A class VII unconventional myosin is required for phagocytosis.  
 A:Reference number: A59249  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: myoI  
 C:Superfamily: myosin motor domain homology  
 /16-676/Domain: myosin motor domain homology <MMO>

Query Match	17.9%	Score 1921;	DB 2;	Length 2357;
Best Local Similarity	25.6%	Pred. No. 6.8e-79;		
Matches 652;	Conservative 402;	Mismatches 721;	Indels 768;	Gaps 81;

```

Qy 62 VDDNASLTLEHGCSIMTNLFOYKRNQIWTYIGSILASVNPFOPIAGLYEPATMEQYSR 121
Db 15 VEDMTTLPLTEESLTLNLMKRYKKKEIYTYGSILVAVNPAE-LIPITADIYKSYPAK 73
Qy 122 HUGELPHIPIANECYRGLMKRHNOCILIGESAGKTESKYLKRLKLSVISOQSL 181
Db 74 SRNLMLPHIPIAASDAFTMMIEGKNOSIIIGHSAGKTESKYLIOYLA----- 124
Qy 182 SLKEKTSQVCERALIESSPIEMAEFNAKYNNNSSRFGFVOLNIOCKNGIOGRIYDCI 241
Db 125 ARTNRHSQVEQMTIVASSPILEAFNAKTRNNNSRFGFPIEQFMRBGHISGARLINYI 184
Qy 242 LSSQNRVVRONGEENRYHIFVALLAGLEHEBEREFLSTPENRYHYNOSGVEDKTIISDQ 301
Db 185 L-EKSRISHQASSERNYHIFYOLLGASDELKELKLGPEPDHYLSQSGCIRIENIN 243
Qy 302 ESFREVITAMDMVQFSKEEVRERYSRLLAGILHGNIEFT-----AGAQVSFKTALGR 355
Db 244 EDFEHAKYAMNYGLPEDKQFIIFSIVSAVLHIGNLKFESKEKTEQGAESSEVSNKDTLKI 303
Qy 356 SAEILGLBPTQUTDLTQKSMFLRGEELITPLVQOAVDSRSLAMALVACCEWYIKYI 415
Db 304 IAGLLSVDPVKLETCLITTRHVLIRGQNFVYPLKVNABEOTRSLALAGVNNMVLVFI 363
Qy 416 NSRI-KXGNEDFSIGLIDIFGFENFEVNFPEOFINIVANEKLOEYENKHIIFSLEOLEYSR 474
Db 364 NSKIHKPOKQNSTFIQVLIDIFGFENFKKSPFEOCFINPAHEKLOQHNOHIFKLEQEBEYER 423
Qy 475 EGLVMEIDWIDNGECLDIEKK-LGLLALINEESHFPQATDSTLLEKLSQHANNHFPY 533
Db 424 EKINSKIYVNDNOECLDIEKRPGLIISLDBEBSFPQATDITLYDKLHTNHEKAPYE 483
Qy 534 KPRVAVNPFVGHYAGEVYDVGRGLIEKKRDTFRDDLMLNRESRDFIYDFEHVSSRN 593
Db 484 KPRRSKNTFVGHYAGEVYDVQGFADKNDKTVYSDLSLSQSKSRFIIELE--VPRP 541

```

```

Db      1532 CEHQSSSLPYPKNF-----VLAALNNGINITYDPATSKMLSEYKYN-----CQSQ 1575
Qy      1478 TKLNATRWSSVIONVDTTKAPIDPTQOLIODIDENKNN-----SVV 1521
Db      1576 QNLKSDKSVSLLENKSTLQA-PTGDVQKVLVSLIKESLYLRNNAKYARALKDVAVSLT 1634
Qy      1522 VEOIYKRNPIRLRYT----- 1535
Db      1635 SLLPFRKNDITITTFKQDENKMFQGLNGKESFPVDHVEILSDVPPQVHPVATLSB 1694
Qy      1536 ----- 1535
Db      1695 PMSPTINNTTPPPPSISDSMSPPPOVGMLEPPPPPSVWGSKPIEIPSLGIPPPPS 1754
Qy      1536 -----HPLHSPPL--LPLPYGDINLNLKDKGYTT----- 1563
Db      1755 SSNSVSPNSBPGSPMKGIPPPPTTISVHSLNSGNSITPPPLPSLTPTLSTPPPISSP 1814
Qy      1564 ----- 1563
Db      1815 PNRSSLRVSMNTSNDGNSDDSKRLTVSPALGTDSQLAQMASTFRSPKRASTLN 1874
Qy      1564 -----IQDEAIKIPNSLQOLESKSD-PIP- 1586
Db      1875 QOQATLKRKAPVDPNTAFYFNKDPKIPESLIEEMAKLSKAIKVFSEIMMW--MGDYPIPK 1932
Qy      1587 -----IIQGILOTGHLRLRDELVCOLIKOTNKVPHGSGVGNLWSMOITLCTGCTPLP 1640
Db      1933 GGTASLIVGSIISRGIEHNEHRLDEIYCOAVROTNNKPKVESAKK--GFELITFLSTTSP 1990
Qy      1641 SRGILK-YLKFHLKR--IREOPG-----TEMEXKALPTYESLKKTKCREPVPSP 1687
Db      1991 SBLDLPFMQKLSRNIAIQSSSPOLASILAVCIKLSHPISPYQO-----RKMGPSA 2044
Qy      1688 DEEALIHQEMSTYVCHGGGCKT-----TINSHTAGVVEKLRGLAM-ED 1736
Db      2045 TEL-----QSFRRSNLENGDISTCKIRFIDOSTKLAKINITYTIRITDVTVCROYGISQO 2098
Qy      1737 SRMFPALFEYN--GHVDKAIESRTVADVLAKEFKLAATSEVGDLPWKFPFLKCYCLDTP 1794
Db      2099 STMGFISAVNEFAGISKVSSETDMTYDVARNEQ--SEEGKE--PYFOVRRRFPID 2151
Qy      1795 NVPK-----DSVEFAFMFEQAHNAVHNG--HHAPAEENLQVLAALRLQYLOGDYTL 1843
Db      2152 DVNKLIDQEHMTDDDICFELTYCOJRDENMKGLYNNAEKOSIIAAILIQILYNGS- 2210
Qy      1844 HAAIPPLEEYVYSLQRLKARISOSTKTFTPCERLEKARTSEFLEGTLRSPRTGSVVRQKYE 1903
Db      2211 -----KLVLTYEVVRQVLP 2224
Qy      1904 BEOMLDMWIKKEVVSARASIIDKMRKFOGNNQOAMAKYM--ALIKEMPGVGSITFDVE 1960
Db      2225 DOJLNSQNIKWMISMESQI-----FELVSQTPRYLKLMTFNLISKSEBPGCTILFNQJ 2278
Qy      1961 CKEGGEPOEIMLVGSADAVSVY--KRGEGRPLEFVFOYEHILSPGALANTYKIVLDER-- 2016
Db      2279 QKEN--PPAKWLAINKKGVSIIPDHTKESKNFWTPOSISNVAF--TDOTFCIMTGNLAK 2333
Qy      2017 --ELLPETSEVDVAKLMKAYIS 2037
Db      2334 PIKQFTTDEHSSIASVYOFPSS 2356

```

## RESULT 5

T30870  
 myosin VIIa - mouse  
 N:Alternate names: Gene shaker-1 protein  
 C:Species: Mus musculus (house mouse)  
 C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 02-Feb-2001  
 C:Accession: T30870, S60346  
 R:Walsh, J.; Mburu, P.; Gibson, F.; Liu, X.; Saw, D.; Steel, K.P.; Brown, S.D.M.  
 submitted to the EMBL Data Library, December 1996

```

A:Description: Mutation analysis of the mouse myosin VIIa deafness gene.
A:Reference number: 220913
A:Accession: T30870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2215 <MAL>
A:Cross-references: EMBL:U81453, NID:g1778381, PID:g1778382, PIDN:AAB40708.1
R:Gibson, F.; Walsh, J.; Mburu, P.; Varela, A.; Brown, K.A.; Antonio, M.; Beisel, K.W.;
Nature 374, 62-64, 1995
A:Title: A type VII myosin encoded by the mouse deafness gene shaker-1.
A:Reference number: S60346; MUID:95174881; PMID:7870172
A:Accession: S60346
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 96-156, 'R', 158-237, 'E', 239-681, 'V', 683-726 <GIB>
A>Note: this sequence variant is from the mouse deafness gene shaker-1.
A:Gene: sh1, MyovIIa, MGI:Myo7a
A:Cross-references: MGI:104501
A:Map position: 7:48.1
C:Function:
A:Description: seems to be involved in auditory transduction
C:Superfamily: myosin motor domain homology
C:Keywords: nucleotide binding, P-loop
F:68-726/Domain: myosin motor domain homology <MMO>
F:96-726/Domain: myosin motor domain homology (fragment) <MMOT>
F:158-165/Region: nucleotide-binding motif A (P-loop)
Query Match      17.2%  Score 1840.5;  DB 2;  Length 2215;
Best Local Similarity 26.3%;  Pred. No. 2.7e-75;
Matches 637;  Conservative 384;  Mismatches 817;  Indels 585;  Gaps 88;
Qy      4  OSTRVWL-RENGQHPSTVNSCAEGIVFRFDYGVQVFFYKOSTIRH-----QKVTAMH 55
Db      6  KQDIYVMMLKSGQEDVPI-----GAVYKLCDSQGIQVVDDEDNHMTSPQNAHIKEMH 60
Qy      56  PTNEGVDDMASLTFLHGGSIWYNI-FOYKRNQIWTYIGSILASVNPQIPLAGLYEPATM 115
Db      61  PTVSHGVEDMIRLQDLNAGILRNRLIRYRDLITYGSLIVAVNPYO-LLSISPEHI 119
Qy      116  EYSRHIGELPPIHFAIANECYRCMKRHNOCLLIGESGAGKSTSTKILKELSVS 175
Db      120  RQYTKKIKGEMPHIFALADNCFYMKRNNDQCIIIGESGAGKSTSTKILQFLAIS 179
Qy      176  QOSSLISLKETSCEERAILLESSPIMEAFNAKATVYNNNSRFKGFVOLNIOCKNIOG 235
Db      180  GGH-----SWIEQVLENTPILEAFNAKATIRNDSSRFKGIYDIHFKRAIIGA 230
Qy      236  RIVDCILSSQNRVVRNGERNYHIFYALLAGLEHEEREEFYLSTPENYHYLNGCYVED 295
Db      221  KIEQYLL-SKSRVCRQADDERHYHVFYCMLEGMNEEEKKXGLGQADYNYLAMGNCITC 289
Qy      296  KTISQESFREYITAMDVMOGSKSEVRSYLLGILGNE-----FITAGANVSFK 350
Db      230  EGRVDSQEVANIRSMKVMLEFTDENWEISKLLAAILMNGILQYARTFENIDACEVLS 349
Qy      351  TALGRSAELIGDPQLTDAIQRSMFLRGEIILPLNVQOAVDSRDSLAMALACCEM 410
Db      350  PSLATRAASILEVNPDDMSCLTSKRLTIRGEIVSTPLSREGALDVRAFAVGIGYRLFTW 409
Qy      411  VIKKINSRI-----KGNEDFKSIGIDIPGFENFVNHFPQFINVANKLQEFYFKHI 464
Db      410  IVEKINAAIYKPPPLEVKNRSRSIGLLDIFGFENFTVNSFEQICINFANHEHLQCFVRHV 469
Qy      465  PSLQOLEYSRGLVVEDIDMDINGECULI-EKLGILLALINEESHFPQADSTLLEKTH 523
Db      470  FTLQEEYVDLSIDWLIHFTDNOALDMINRPNVNSLIDESKFPKGTATYLAHLKN 529
Qy      524  SOHANHHFYVVRKVA-VNPFGVKHVAGEVQVDVRGILTEKNRDTFFDDLLNLRESRPFI 582
Db      530  SOHKLNANYVPPKNSHEQFGINHPAGVYVYESQFLKRNDDTLHGDIILQVHSSRNKFV 589
Qy      583  YDLFEHVSSRNQDTLKGSKRHR--PTVSSQFKVDSLHSLMATLSSNPFVRCIKPMN 640

```

```

Db      590 KOIF-----CADVAMGAETRKSPITSSQFK-RSLELMKRTIGACQPFVRCIKPBE 640
Qy      641 OKMPQPOAVVNLQRLYSQMLETRIRIRAGAVARPPQDFYKRVKVMRLNA---LPRD 697
Db      641 FKXPMFLPDLHLTCVROLRYSQMMETIRIRAGVPIKRSFVEFERKRVLLPGKPAVYKQD 700
Qy      698 VNGKTSLLQLYDASNSEWQGLTKVFLRESLHOKLEKREBEVS-----742
Db      701 LMGTOQMAEAVLGTHDWDQIGTKIFLKDHMDLLEVERDKAIDRVILLQKVRGPD 760
Qy      743 -----HAMVIRAH-----VLGFL-----ARKOYKRVLYCVVIT 771
Db      761 RSNPLRLKSAATLIQRHMGHCHCRKUYELIRLGLFLQALHRSRKLHKYRLARORILEF 820
Qy      772 OKNVAPFLRRRFLHKKAAIVFOKOLRGQIARVYROLAE--KREOEKKKQBEBEKK 829
Db      821 QARCBAYLVKRAFRHRLMAVITVQAVAKMIALRHLRLVAYQRRLERMRRLAEBEKL 880
Qy      830 KRE-----EEREREREREREAELRAQOEETRKQOELEALQSKQEAELTELEKQK-- 881
Db      881 KREMAKKAKEBAERHGERLAQLABEAE-----RELKEKEARRKKELEQMEKARHE 935
Qy      882 ---ENKQVEEIL-----RLEKEIEILOQMEQOEELSLTEASLOKLOERRDQ 924
Db      936 PINHSDMDVKMGFLGTSGSLPQEGQAPSGFEDLERGR-----974
Qy      925 ELRLREBEACRAQEBLESLNDFEIDECRNTERLSGSEFSSSELAESACEKNFNFS 984
Db      975 ---RENAVEVDNALPLP---DEDEDESEYKFAFAATVYFOGTTTHS-YTRRP--LK 1023
Qy      985 QP--YPEREVD-----GFEADDAFKQSDPNSEHGHSDORTSGIRTSDDS 1028
Db      1024 QPLVYHDDGDDQALAAVITILRFMG-----DLPERKYN-----TAMSG 1064
Qy      1029 SEEDPYM-----NDTVVPTSPADST--VLLAPVQDSGSLN 1064
Db      1065 SEKIPMTKIYETLGKTKYKRELQALQEGEQTOLPEQOKTSVRHLLVHLLTKKSKLITE 1124
Qy      1065 S-----SGSESTYCMQONAGDLPSPD-----GDVYDQDDYEDGATIS 1102
Db      1125 EYTKRLNDGESTVQOQNSMLEDRPTSLEKHLHLLIGNGILRALREIYQISQK--LTH 1181
Qy      1103 GSVYTFNSYGSQMSPDYRCSYGTYNSS-----GAYRFSSEGAOSSPED 1146
Db      1182 NNS---KSYARGMILVSLC-VGCRAPSEKPYKILANFIHGPRPGAPRCEERLRTFPN 1237
Qy      1147 SEEDPSRPTDDELSYRDSVYSCVTLRFHSLYMKGLMN-----SMKRRMC--- 1196
Db      1238 G-----TRTOPPSMLELQATKSKKPIMLP---VTFMDGTTKLLTDSATYARELCNALA 1288
Qy      1197 ---VLKD-----ETPLMFRSKQELAKQGMH-----KKGGSSTLSRRMMKRW 1237
Db      1289 DKISLDRFGSLYLALPKVSSLSGSDHVDALISQCEYAKEQQA---QERNAWML 1344
Qy      1238 FVLROSKLAMYFENDSEELKGVY-----EVRTAKB--IINDTTKENGIDY--- 1280
Db      1345 F-FRKEVTFPMNPSDDNATNLIIQOVVGVKFGYRCEKEDDLAELASQOYFVDYSGE 1403
Qy      1281 INADRTFLHIAE-----SP-EDASQMFVLSQVH---ASTDQEIQEMHDEQANPON 1327
Db      1404 MLERLRLSLVPTIYIPDREITPLNLEKMAQLALAHKKGIVQORRDSQKVEDVNVYAR 1463
Qy      1328 AVGTLDVGLIDVCSADSPDRPNSFYIIRANRVLEHGNATPREMHMWTLLQSKQDTRV 1387
Db      1464 FKMPLFSRFYEAYKFSPPPKSDVIYVNMVTGYFVEBQOV---LLEISPEELWA 1518
Qy      1388 --EGGEFYI-----RGWLHKEVKNSPKMS--SLTKLKRWFVLTNSID 1426
Db      1519 VSSSRRCRVLLSLGCGLCATCQSGARAGLITRAGPSCPCMSCRGTGMMASTFLATIKBD 1578
Qy      1427 YKSKSEKNAKLGTLVNLSCVVPPEDEKIFKETGHW-----NTVYGRKHGCRLYTK-- 1479

```

```

Db      1579 EYFTSSNAEDIRDLVTFLL-----EGLKRSKRYVALQDNPNPAGESGFLSFAKGD 1631
Qy      1480 --LNEAT-----RMSGVIONVDTKAPIDT-----PTOOLI----- 1509
Db      1632 LIIDHDITGEQVYMSGNANGINERTKQGRDPTTCQVYVMPVTIIPREIYALVYMTTPDOR 1691
Qy      1510 QDI-----KENCINSNDVVEQIYKRNPIILRYTHNP 1538
Db      1692 QDVVRLQLRTAEVBRAPKPYTLEEFSSYDYFRPPPKHTLSNVMSYKANGKRLMSHTBEP 1751
Qy      1539 LHSPLPLPYGDINDLNLKDKGYTTLODEAKIKNSL-----QOLBSMSPDIPITQ 1589
Db      1752 LKQALK-----KILGSEBELSQEACMAFAVALKYNGDYPSKMRSSVNE---LTD 1797
Qy      1590 GILQTHGLRLPRLBELYCOLIKQ--TNKVPPRGSVGNLYSQOITLCSCTFLPSRGLIKY 1647
Db      1798 QIFEMALKAEPLKDBAYQIILKOLDNHIRISEERG---WELMLCTGLFPSPSILLPH 1853
Qy      1648 LKFLKRLIREQFPGTEMEKVALFTYESLKK--TKCREFVPSRDEIALIHR-QEMTSTV 1703
Db      1854 VQRFLOS-RKQCP-----LAIDCLQRLQKALRNGSRKYPHLVVEVAIQHKTQIPIKV 1906
Qy      1704 YCHGGGCKITTINGHTTAGEVVEKILRGLAMEDSRNMPAL-----PEY 1746
Db      1907 YFPDDTDAFVEBSSTKAKDFCONIASRLILKSEBG-FSLPVKIADKVISVENDPFPDF 1965
Qy      1747 NGHYDKATESRTVADVLAKEPKLAATSEVGDLPMKFPF--KLVCFLDTDNVP-KDSV-E 1802
Db      1966 VRHLTDWIKKARPIIDGI-----VSLTYQVFPMKLM---TTVPQKDPAD 2010
Qy      1803 FAFMEQNAEAVIHGHNPAPENILOVALA--RLQYLOGDYTLMAAIPLEBYVSLQRLK 1860
Db      2011 SIFHYQELPKYLGHYKCTHEBVLQGLALYRVXF-BEDHSYFSPISPKL----- 2059
Qy      1861 ARISQSTFTFPCRLEKRTISFLEGTIRSRFRGSVYKQVBEQOHLDMITKEBVSSAR 1920
Db      2060 -----LR-----ELVPQDLIRVSPDDM-----K 2078
Qy      1921 ASIIDKMRKFGQMNQOAMAKYMLIKEMPGVSTLPDV-ECKEGFPOELMIGVSADAV 1979
Db      2079 RSIYAVFNKNAKSKKEBKALFLLKILFMPFFGSAFEBVKQKTTEBNFELILIAINKGV 2138
Qy      1980 SVYKREGRPLEVFOYEHILSFGAPLANTYKIV-----VDERELLFETSEVVDVAKLMA 2034
Db      2139 SLIDRPTDOLLTHPTISWMS--GNTYHITIGNLVRSKLLCETSLGKYMDLLTS 2196
Qy      2035 YISMIVKKRYSTTSASQSSR 2057
Db      2197 YISQWL-----TAMSKQRNSR 2212

RESULT 6
A:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 02-Feb-2001
C:Accession: A59255; S71932; S60263
R:Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M.
Genomics 36, 440-448, 1996
A>Title: Molecular cloning and domain structure of human myosin-VIIa, the gene product d
A:Accession: A59255; MUID:97036886; PMID:8884267
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2175 <CHB>
A:Cross-References: GB:U55208; NID:G1613787; PIDN:AAC50927.1; PID:G1613788
A:Experimental source: tissue type testis
R:Well, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S71932
A:Accession: S71932
A:Molecule type: mRNA
A:Residues: 96-334,'T',336-370,'N',372-388,'VTG',392-469,'F',471-563 <WEI>

```

A:Cross-references: EMBL:U17180; NID:G1235669  
 A>Note: the sequence is revised in GenBank entry HSU39226, release 114, (PIDN:AA03679.1)  
 R:Wells, D.; Blanchard, S.; Kaplan, J.; Gillford, P.; Gibson, F.; Welsh, J.; Mburu, P.; V  
 Plet, D.; Munich, A.; Steel, K.P.; Brown, S.D.M.; Petic, C.  
 Nature 374, 60-61, 1995  
 A>Title: Defective myosin VIIA gene responsible forusher syndrome type 1B.  
 A:Reference number: S60263; MUID:95174880; PMID:7870171  
 A:Accession: S60263  
 A:Molecule type: mRNA  
 A:Residues: 96-157,199-255 <MEM>  
 A:Cross-references: EMBL:U17180; NID:G1235669  
 A>Note: the sequence is revised in GenBank entry HSU39226, release 114, (PIDN:AA03679.1)  
 C:Comment: A alternate transcript from the same gene is shown in PIR:A59257.  
 C:Genetics:  
 A:Gene: MYO7a; USH1B  
 A:Map position: 11q13.5  
 C:Superfamily: myosin motor domain homology  
 C:Keyword: actin binding; ATP; nucleotide binding; P-loop  
 F:96-563/domain: myosin motor domain homology (fragment) <MMOT>  
 F:158-165/Region: nucleotide-binding motif A (P-loop)

Query Match 17.1%; Score 1833; DB 2; Length 2175;  
 Best Local Similarity 26.2%; Pred. No. 5.8e-75;  
 Matches 632; Conservative 368; Mismatches 802; Indels 614; Gaps 85;

QY 4 QGTRVWL-RENGQHPSTVNSCAEGIVFRTDYGVFTYKQSTITTH-----QKVTAMH 55  
 DB 6 QGDHVMMDLRSGGEFDPVI-----GAVVTKLDCSQGVQVNDENEMISQNNATHKPH 60  
 QY 56 PTNEGEVDMASTELHGSIMYNLFQRYKNOIWTYIGSILASVNPYQPIAGLYPATM 115  
 DB 61 PTNVHGEDMIRLGDENAGILRNLLIRYDHLIYTGSGILVAVPYQ-LASTYSPH 119  
 QY 116 EGYSRHLLGLPHIFAIANEYRCYLMKRDNOCILIKGSGAGKTESTLILKPLSVS 175  
 DB 120 RQTNKKIGEMPHITAIANCTYFNMKRUSRDCCIIISGSGAGKTESTLILQPLAAS 179  
 QY 176 QQSLELSEKETSCEVERALLESSPIIEAFGNACTVNNNSRFGKQVQANI CQKGNIQG 235  
 DB 180 GQH-----SMIQVLEATPILEAFGNACTINNDNSRFGKQYIDHFNKRGALIEGA 230  
 QY 236 RIYDCLISQNRVVRONGERNYHIFYALLAGLEHEEREFPYSTENYHYLNQSCVED 295  
 DB 231 KIEQYVL-EKSRVCRQALDERNHYHFCMLEGMSBQKKLGLGQASDVNYLGMNCITC 289  
 QY 296 KTSIDQSPREVLTADVWQFSKEEVRVSRLLAGILHGNIE-----FITGCAQVSRK 350  
 DB 290 EGRVDSQBYANINSAKVLMTDTENWEISKLAAILHGNLQYEARTENIDACEVLS 349  
 QY 351 TALGRSAELGLDPTQLTALTORSMFLGEEILTPLANVQAADVSRDSLALAYACCFEM 410  
 DB 350 PSLATASLELVNPPMLMSCLTSRTLITGETVSTLSRQALADVADAPKGIYGLFLFW 409  
 QY 411 VIKKISRTI-KGNDEPK---SIGLIDIFGFENFEVNHPEQFNIVYANEKLOEYNNKII 464  
 DB 410 IYDKINAAAYKPPSQVKNRSRISIGLIDIFGFENFVANSPEQICINFANEHILQGFVRYH 469  
 QY 465 FSLQLEYSRREGVWEDIMDNGECLDI-EKKGLALLINEESHPPQATSTLLEKH 523  
 DB 470 LKLEQSEYIDESTIDMLHIEFTDQDADMLANKPMIISLIDEESEFPKGTDTMLHKN 529  
 QY 524 SQHANHHFYVKEPVAVNN---FGVGHYAGEVQYDVRGILEKNRDTFRDILLNLIESRF 579  
 DB 530 SQHKLMANYIPK---NNHETQGINHPAGIYVEYEGFLEKNRDLTHDDIQLVHSSSN 586  
 QY 580 DFIYDLFEHVSNNODTLCKSGKRR--PTVSSQFVDSLHSLMATLSSNPFVRCIK 637  
 DB 587 KPIKQIF-----QADVAMGAETRKRSPTLSSQPK-RSLLEMLMTLGAQCPFFVRCIK 637  
 QY 638 PNNQKMPDQDAVVLNQLRYSGMLFTVIRKAGVAVRRPFODFYKRYKVLKRNLA--L 694  
 DB 638 PNEFKKMLPDRHLQVROLRYSGMELTIRIRAGYPIRYSFVEFERYVLLPGVAPAYK 697

QY 695 PEDVKGCTSLQLDYDASNSSEWOLGKTIVFLRESLEOKLEKREEREESHAAVIRAHVVG 754  
 DB 698 QGDINGTCQRMAMENAVLGHDDWQIDKTKILFKDHHMLLEVERDAIYDRVILLQKVRIG 757  
 QY 755 FLARKYRKVLVCVVIQ-----KNY----- 775  
 DB 758 FDRSNFLTLKNAATLQHNMRGNCRKNYGLMRGSLRLQALHSRKLHQYRLARORI 817  
 QY 776 -----RAFLRRBFLLKKAIVFOKQRLGQIARRVROLARK--REDEKKKQEE 826  
 DB 818 IQFQRCRYLVKQAFRRHLMAVLTVQAVARQMILRRHLHQRARAYLRLBEKKQALAE 877  
 QY 827 EKKKEEEREREREREREALEAQOEEETRQOELALQSQKEALTRLEKQENKQV 886  
 DB 878 EKLREKMSAK-----KAKEAEKQGERLAQLABEDAERL----- 913  
 QY 887 EBITLKEIQLQMKQOELSTLEASIQKIQERDQELRLLEBACRAQEFLESINF 946  
 DB 914 -----KEKEAARAKE-----LLEQERARHBPVN-----HS 940  
 QY 947 DEIDCVNRNIEKSLGGSEPSSELAESACEKPNFNFGPYPEEEDVGFEADDDAFKDS 1006  
 DB 941 DNVDMKFGFL-CTSGGLPGQEGQAPSGFEDL-----ERGRREWEEDLDALAPL 988  
 QY 1007 PNPSEHGHSDQRTSGIRTSIDSEEDPYMNDTVFPTSPADSTVLLAPSVQDSGLHNS 1066  
 DB 989 PDEDEEDLSEYKFAFAFAT-----YFGGT--TTHSYTRPLKQP-----LLYHDD 1031  
 QY 1067 GSESTYCNP-----QNAQDLPSPDGDDYDQDD-----YEDCAITSGSSVFS 1109  
 DB 1032 EBDQALAAVLWITILRFMGDDLEP--KYHTAMSDSEKIPVWTKIYE--TLGKQTYKR 1085  
 QY 1110 NSYGSQWSPDYCSVGTNNSSGAYRFSSEGAQSPDESE-----DFPSRDTDELSY 1163  
 DB 1086 ELQALQGEBAQLPBGQKSSVYRKLVLTLKKSRLTEVYTKLHDESESTYQGSMLD 1145  
 QY 1164 R-----RDSVYSCVTLLPYHSEFLYMKGLMNSMKRW----- 1195  
 DB 1146 RFTSNLEKLHFIENGILRPLARDEIYQISQLTHN-----PSKSYARGMILVSLCV 1199  
 QY 1196 -CVLKDEFTFLMRSK-----QEALQGMHLKKGGSSTLSRRMKKMPVLR 1241  
 DB 1200 GGFABSEKVKYLRNFIHGPGPYAPYCEERLRRFTV-----SGT--RTQPSMLBQ 1250  
 QY 1242 QSK-----LMVFENSEEKLKGTVEVRTAKIID-----NTKENGIDIIA----- 1283  
 DB 1251 ATKSKKPIMLPVTFMDDGTTKTLITDATTAKELCALADKISLKRFGSLYIALFDKYS 1310  
 QY 1284 -----DRTFHLIASESPDASQ-----W-----FSVLSQVASTDOEIQ----- 1316  
 DB 1311 SLGSGSDHVMDAISQCEQYAKQGAQERNAPRLPFRKEVFTPMHSPEEDVATNLIIYQ 1370  
 QY 1317 -----EMHDEQANQNAVGLDVG-----LIDSVCASDSDPRNSVVIITA 1357  
 DB 1371 VVRGVKFESEYRCEKEDDLAELASQOYFVDYSGEMILRLNLV-PTYLPDR-----ITP 1424  
 QY 1358 NRVLHCNADPTEPM-----HWITLQOR----- 1380  
 DB 1425 LKTLERKQALAAHKQIYAQRRTDAQKVEDVSYARFKMPLFSFYENYKSGSGL 1484  
 QY 1381 SKGDT-----RVEGQEFIVRGMLHKEVKSPPKNSLKLKKRWFLTHNSLDYYS 1430  
 DB 1485 PKNDVIVAVNMTGVYFVDEQOVLLELSEFPEIMAVSSRGAKTAPSTLATIKDEYTF 1544  
 QY 1431 SEKNAKLGITLVNLSLCSVPPDEKIFPETGYW-----NTVYGRKHCYRLYTK--LL 1481  
 DB 1545 TSSNEDIRDLVAVFL-----EGLRRKSYVVALQDNPPPAGESEGLSFAKGLLITL 1597  
 QY 1482 NEAT-----RWSVIGNVTDTKA--PIDT-----PTQOLI-----ODI- 1512  
 DB 1598 DHDIOEQVMSGWAGINERTKRGDFPTDSYYVMPTVMPRELRVLTMTPDQRODV 1657  
 QY 1513 -----KENCINSDVVEQIYKRNPIRLRYTHHPLHSP 1542

Db 1658 RLLOQTABEVRAPKPYLTLEESYDYFRPPPKITTSRWVWSKARGDRMSHTRBPKQA 1717  
1543 ILPLPGDINLN-----LLKDKGYTTLQUBAIFKINSLOQLESMDPIPIIQGIQ 1593  
Db 1718 LAKKLGLSELSEGBACLATIAVLKTVWG-----DYPSKRTSRVNELTDO-----IFEGPIK 1767  
Qy 1594 TCHDLRLPLDELVCQLIKQ--TNKVPSPSGVNLVSMQILTCLSCFLPSRGLKYLKPH 1651  
Db 1768 A-----BPLKDEAVVQLKQLTDNHIRYSEERG---WELLMLCTGFPPSNILPLPHVQPF 1819  
Qy 1652 LKRIHQPFQTEMEKALTTYESLKK---TKCEFPSPDEIFALIR--QENSTSYVCGG 1707  
Db 1820 LGS-RKQCP-----LAIDCLQLQKALRNGSKYPPHLVEVEALQHKTTQIIFHKYFED 1872  
Qy 1708 GGSCKITINSHTTAGSEVVEKLIRGLAMEDSRMNFAL-----FEYNCHV 1750  
Db 1873 DTDEAFEVSSSTAKQDFCONIATRLILKSSG--FSLPVKIKADKVLSPENDEFFDVRHL 1931  
Qy 1751 DKAIESRTVADVLAKEFKLAATSEVGDLPWKYF--KLVCPLDTDNP-KDSV-EFAFM 1806  
Db 1932 TWMIKARPIKQI-----VPSLTQVFPFKKLM---TTTVPKQPMADSIH 1976  
Qy 1807 FEOAHAVIHGHHPAEENLQVLAAL--RLQYVJGDYTLHAIPLPEEYVSLQRLKARIS 1864  
Db 1977 YVOELPKYLRGYHKCTREEVLTQGLALIVYKF-EEDKSYFPSPKPL----- 2021  
Qy 1865 QSTKFTPTPERLEKRTSFLGTLRSFRTGSVVRQKVEEQMLDWIKIEVSSARASTI 1924  
Db 2022 -----LR-----ELVPDLIRQVSPDM-----KRSTV 2044  
Qy 1925 DKARKFQGNQEQAMAKYALIKEMPGYSTLPDYVECKEGFPQELMLGVSADAVSVYK 1984  
Db 2045 AEFNKAGKSGKEAKLAFKLIFKWPFPSSAPFR-QTEBPNPEILLIANKGVSLIDP 2103  
Qy 1985 GEGRPLEVPQYEHILSFGAPLANTYKIV-----VDERILLFETSEVVDAKIMKAYISMT 2039  
Db 2104 KTKDILTTHTPFTISIMSS--GNTYFHTIGMLVRGSKLLCETSLGYKMDLLTSYSIQM 2161  
Qy 2040 VKKRVSTRSASQGS 2055  
Db 2162 L-----TMSKORGS 2171  
RESULT 7  
T25888  
hypothetical protein T10H10.1 - Caenorhabditis elegans  
C1Specie: Caenorhabditis elegans  
C1Date: 15-Oct-1999 #sequence\_rev15on 15-Oct-1999 #text\_change 08-Sep-2000  
C1Accession: T25888  
R1Nelson, J.; Langston, Y.  
submitted to the EMBL Data Library, December 1996  
A1Description: The sequence of C. elegans cosmid T10H10.  
A1Reference number: Z20105  
A1Accession: T25888  
A1Accession preliminary; translated from GB/EMBL/DBJ  
A1Molecule type: DNA  
A1Residues: 1-2098 <NUL>  
A1Cross-references: EMBL:U0848, PIDN:ABJ7988.1, GSPDB:GN00028, CESP:T10H10.1  
A1Experimental source: strain Bristol N2; clone T10H10  
C1Genetic:  
A1Gene: CESP:T10H10.1  
A1Map position: X  
A1Intons: 6/3; 43/3; 92/3; 154/2; 226/1; 280/3; 417/2; 472/3; 812/3; 853/3; 887/3; 978/  
/3  
C1Superfamily: myosin motor domain homology  
F165-720/Domain: myosin motor domain homology <MIO>  
Query Match 16.0%; Score 1716; DH 2; Length 2098;  
Best local similarity 25.4%; Pred. No. 1.1e-69;  
Matches 602; Conservative 364; Mismatches 806; Indels 596; Gaps 77;  
Qy 4 OGTRVWLRENGHFPSTVNSCAGIVVFRTDQGV-----FTYKOSTITHQKVTANHPN 58

Db 6 KGDFTWI-EPGK-----TEGSIPIGARVIDQHGRLKVIDIDGNEQMLSADRVRMLMPTS 60  
Qy 59 BEGVDDMASLBLEHSGSITMYNLFORKYKNOIWTYGISILASVNPQPIAGLEVAPATMBOY 118  
Db 61 VQGVDDMQLGDFHESALIRNLFIYRKLIYATYGSILIAVNPMDIA-IYTADEIRMY 119  
Qy 119 SRRLGELPRPIFAIANECYRCMLWRHNOGILIKGSEAGKTESTKLILKFLSVISQGS 178  
Db 120 KRKRIGELPRPIFAIADNAVYTNMRREKNGQVVISGSGAGKTESTKLVLEFLATISQGH 179  
Qy 179 LELSLKETSCEYERILSSPIMEAFGAKTYVYNNSSRFKFPQVLANICQKNTIGGATV 238  
Db 180 -----SWIEQVLANPVLBAFGNAKTIKINDNSRFQYIDVHNEGSGIEGAKIE 230  
Qy 239 DCILSSONRVVRONGERNYHIFVYLLAGLEHEERBEYLTSPENYHILNOSGCEVDXTI 298  
Db 231 QYLL-EKSRIVYQSENERNYHIFYCLLGLSREKSELGTADYTYLIGKITYIAEGR 289  
Qy 239 SPOESFEVITAMDVQPSKEVREVSRLAGILHNIER-----ITAGNAVSPKTA 353  
Db 230 DDAADLAEIRSAMRYLMIHQEIGSIFKLASLHIGNIRFRONTNDMESVDVADPSTL 349  
Qy 334 GSAELGLDPTQLTDALTQSMFLRGEILTPLVNQAQVDSRSLAMALYACSEBVIK 413  
Db 350 VRIATLQLHQNLDAITTKSLVTRERVISRLNGQAQVADALAKAIVGKLFHIVR 409  
Qy 414 KINSRI-KGNDFK-SIGILIDFGFENEVNHFPQFINVYNEKQOEXFNHGISLEOLE 471  
Db 410 RVNDIIVYPSQSRRTSILIDIFGENESESFEOLCINPNETLOQFPVHVHFMEOKE 469  
Qy 472 YSRGELVEMDIDWIDNGECLDLI-EKKGLALALINEESHFPQATSTLLEKLHSHOHANNH 530  
Db 470 YDEEHINRHIKFVNDQATVDLIAQRPLNLSLIDEBEIPKRGDKTMLKLSHNGNE 529  
Qy 531 FYVKRRVAVNN-FGYKHYAGVQYDVRCILGILEKNDTPDDILNLRBSRPFYDLPFHV 589  
Db 530 LYLQKSELQRAFGVTHAGVVFYNTGRFLEKNDRSFADISLVLSISSKMPFLALFPDI 589  
Qy 590 SSRNNQDILKSGSKHRRPFTVSSQPFVDSLHSLMALTSSNPPFVACIIPNMQMDQDPQ 649  
Db 590 ----EYDT-----SSRKATVGNQFR-KSLQOLMSQLTGHFFIKCIIPENKRALVMDR 640  
Qy 650 AVVLNLQRYSGMLETVRIKAGYAVRRPQDFYKRYKLYMENTALPE---DYRGKTSIL 706  
Db 641 DLVLQALYSGMETIKIRBSGYPIRHDYIPVFPYRVLVSSIQCPVARIIDLHDAKKIC 700  
Qy 707 QLYDASNEBQGLTKTVFLRSLEQLEKREBEVSHAMVIRAHVIGFLARKQRYKLY 766  
Db 701 HMLGTNADYQGLTKTVFLKQGHDLVLE-----QEVYRILKD 737  
Qy 767 CVVITQRYRAFLRRFLHLKKAIVQKQRLGQIARVYRQQLAEKREBEKKQOEBE 826  
Db 738 KAIVQKVRRLVAKDEKQKQAAVTTQTAMRGDQKRRYQIIS----- 783  
Qy 827 EKKKKEBERERERRERAEELRAQOEBETRQOELALQSKQKALTRLEKQKENQY 886  
Db 784 -----GFRDLQVLRRO-----LV 798  
Qy 887 BEILBLEKEIEDLQMKQOELSLTEASLQKQERRDOELRLLEBACRAQEFLESJNF 946  
Db 799 SHYQTLRTIITQFO-----AVCRGSLVR---RQVEKRIKRGKAPLTVESSTASVIS 847  
Qy 947 DEIDECVANIERSLSGSEFSSSELAESACEKPNPNFQPIYEEBVDSGFEDDAFQDS 1006  
Db 848 DSHEELVHL-----FDF---LPSDKDSGNB-----NDS 874  
Qy 1007 PNPSEHG-HSDQRTSGIRTSDDSEEDPYMD----- 1037  
Db 875 ADSRRGYSRLHITSFPVNPANIPRVDSYVEDLSKYQPKAAATFPQAQATATHTVKKPL 934  
Qy 1038 ---TVVPTSPSADSTVLLA--PSVQDSGLNHSNGSBSTYCPONAGDLPSPDGDYDYO 1092



```

Db 935 KVALHTHTPEPQAOLALTANTTILRENGDLADYKPG-----STNGSEVYDK 960
Qy 1093 -----DVEDEAITS--GSSVTFPSNSYGSQMSPDYRCVGTYNSSGAY 1133
Db 981 TPVMIKLYATLGKFFSAHDLSEYGAATLKGMKRKILSMTRKKGKINSDDPS 1040
Qy 1134 RPSSEGAQSSFEDEEDFDSRPTDDELST-----RDSVYS-CVTLPYHPSL 1181
Db 1041 SISDSYSSFFNAML-E-NKPMTSLDLYHIIIGILREDBLRDEIYCOQLKSNPNSKL 1098
Qy 1182 YMGGLMNSMKRW-----CVLKDETFLMFRSKQOALKQGMHKKGSSSTLSR--- 1230
Db 1099 -----SARGMWILSLCVCGFAPSERIFKTLF-----CFIRBGAGTYGKYIIE 1143
Qy 1231 -----RMWKRWFVLROSK-----LMYFENDESEKLT---KGVEVTRAKEII 1269
Db 1144 DLRRTQVNTGRQOPPSYVELQANKSQKPYVLAVTFMDSGVKTLCDASATTAELCKOLA 1203
Qy 1270 DNTTKENGIDIIADRTFHLIASPESASQWFSVLQVHA-STDQETQE----- 1317
Db 1204 EKVGLTNSFGFSLYIALFDKVSLSGCTDHVMDAISQCEQYAKEQGRQERNAPWRLFPK 1263
Qy 1318 -----WDEGANP-----QNAV-----CTLDVGLI 1337
Db 1264 EIFSPWMDPRDDVSTNLIYQVIRGIRYGEYRCDEDEIALAICAOYYIDEETMDVNL 1323
Qy 1338 DSVCASDSPD----- 1347
Db 1324 ENNLPSYLPDPFEMSGEMALEKMTQITIMQYRKKTGRRLPSQIEVENVSVAKTWPRL 1383
Qy 1348 -----RNSFYITAN-----RVLHCNADTPREMHHTITLQSRSG 1383
Db 1384 FSRFYEALKEAGPPLPNEVYIIANWTVVVDREHVMLEFSFPEISTAYYKGRKSTT 1443
Qy 1384 DT-----RVSEGEFIVGMLHKEVNSPKM--SSLXKKRFFVLTNHSLDVYKSEKN--- 1434
Db 1444 DTCTVTVVGDDEYTFQSPNADITNLJWFLEGLKGRSRLV---AISOQKDEKONPL 1499
Qy 1435 ALKLGTLVL-----NSLC--SVVAPDEKIFKETGYMNTVYGRKHCVRLYT--KLHNE 1483
Db 1500 EFEEKGDLILVNFNTGNTLILTESVVGKENS-----RTCLFGLIRAEVNYVLPPLVLP 1551
Qy 1484 ATWSSSVIQNTVDTKAPIDTPTQOL-----IODIKENCLNSDV-----VBO 1524
Db 1552 SKNTLQIFPKMDLSLDLFNNKQVTVVDYNAEPYLTLENPAEDNFNSQVAVSGQISLMT 1611
Qy 1525 IYKRNPL---RYTHHRLHSLPLPYGDNINLMLD--KGYTTLDEDAIKTNSIQOE 1579
Db 1612 LRKESQIECWRSREHIDQPLK-----KLNGREDACRG-----ALIFPAI--WK 1656
Qy 1580 SMSDP-----IPITQILQGTGHLRLPRLDELVCOLIKOTNKVPHPGSVGNLYSMQIL 1631
Db 1657 YMDDEBSKSRGLQTHLDHIFKLPISMEALRDELQYQVQQLT--NPSIMSERGWEILL 1714
Qy 1632 TCISCTPLPSRGILKLYLKLKRIKIREOPCTEKEXAL--FTYESLKTKCEFPVPSBD 1688
Db 1715 WMATGFAPSAALAKESHEFLK-----SRPHIALDCQNMOKLAKGSGSKYPPHLY 1766
Qy 1689 EIALALHR-OEMSTVYCHGSGCKTITNSHTAGVEVEKTLIGLAMEBSRMNFALEFXN 1747
Db 1767 EYVAIQHTKTQIHNKVPFNDTDEALEVDSATAPRCHKIGYRLGKSS--DEFSILF--- 1822
Qy 1748 GHVDAKIESRTVADVLAKEFKLAATSEVDDLPMKEFYKLYCFLDTDNPVKOSVEPAFNF 1807
Db 1823 -----VKIKOKVLA-----VPES--EFFPY 1841
Qy 1808 EOAHEAVIHGHAPPE-----NLOYLAALRLQYLOQDYTLHAALPRLBEVYSIORLKA 1861
Db 1842 VRLSLDMVYTNHATQDATWIPINYQYFWRKLMY---NFVAGADPOADIIRHYHQ--- 1894
Qy 1862 RISQSTKTPPCRLKERTSFLSEGLRBSFRGSGVVRQKVEBEOQLDMWIKEE----- 1915
Db 1895 ---ESQYLLGYNHKTKNVIELAALILRSMTKDGKNAPLAQIPQLLDEIIPKDSLKMS 1951

```

```

Qy 1916 VSSASASIIDKRRKEQGNQOQAMAKMALIKEMPGYSTLFDV-ECKEGFPQOELMLGV 1974
Db 1952 ASEWKTISMAYARIEHLKSDQAKIEPLNYICRMWTFPSGAFPPVQSYDNLNPDRLILAI 2011
Qy 1975 SADAVSVYKRGGRPLEVFOYHILISFGAPLANTY-KIVY-----DERELLFETSEV 2026
Db 2012 NOTGVNIYHLDYKNLLVQYPPFVNIQNTS--GNTYFNMTVGMKMGNGKLLDITVGY 2069
Qy 2027 DVAKLMKAYISMIVKRY---STRSAS 2051
Db 2070 KMDDLTSTIISLISQNNHPSKTRERA 2057

```

## RESULT 8

```

A:59257
myosin VIIa, short form - human
C:Species: Homo sapiens (man)
C:date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 01-Mar-2002
C/Accession: A59257
R:Chen, Z.Y.; Hasen, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M.
Genomics 36, 440-448, 1996
A:title: Molecular cloning and domain structure of human myosin-VIIa, the gene product d
A:reference number: A59255; MUID:97038686; PMID:884267
A:Accession: A59257
A:status: preliminary; not compared with conceptual translation
A:molecule type: mRNA
A:Residues: 1-1203 <CHR>
A:cross-references: GB:U55209; NID:g1613789; PIDN:AAC50722.1; PID:g1613790
A:experimental source: tissue type testis
C:comment: A alternate transcript from the same gene is shown in PIR:A59255.
C:genetics:
A:gene: MYO7a
A:Map position: 11q13.5
C:superfamily: myosin heavy chain 95f; myosin motor domain homology
F:88-729/Domain: myosin motor domain homology <MNO>

```

```

Query Match 15.9%; Score 1702.5; DB 2; Length 1203;
Best Local Similarity 37.1%; Pred. No. 2.1e-69;
Matches 382; Conservative 196; Mismatches 313; Indels 139; Gaps 24;

```

```

Qy 4 QSTRVWL-RENGQRPSTVNSGAEIVYPRDYGQVFPYTKOSTTH-----QKVTAMH 55
Db 6 QGDHYWMDLRSGQEDVDPI-----GAVVYKLCDSGVQVVDDEDEHMSIPQNAHIKPMH 60
Qy 56 PTNEGVDMAELTEHSGSTVYNLFORYKRNQITVYGSILASVNPQPIAGLEPATM 115
Db 61 PTVSHGVEDMIRLQDLEAGILRNLLIRYRDLITTYGSLVAVNPYQ-LISTISPEHI 119
Qy 116 EQYSRRHLGELPPHIFAIANBECYRCIAMRHDNQCILINGESGAGKTESKTLKEFLSYIS 175
Db 120 RQYTNKIKGEMPHIFALADNCYFMKKNRSHQCCIIIGESGAGKTESKTLIGFLAIIIS 179
Qy 176 QQSLELSLKEXTSCYERAILBESSPIWEAFGNKATVYNNNSRFGKFFVQLNQCQKNTQGG 235
Db 180 GQH-----SWIEQOVLEATPILEAFGNKATIRNNSSRFGKYIIDIHFNRKAIGIA 230
Qy 226 RIVDCILSNRVNQNGERNYHIFVALLAGLBEHEBEFYLSPENYHYLVNQSGVED 295
Db 231 KLEQYLL-EKSRVCKQALDERNYHVFYCMLEMSMSDQKKLGLGASDYNYLAMNCITC 289
Qy 236 KTISQESFREYITAMDVQFSKEVERVSRILAGILHAGNIE-----FITAGAQVAFK 350
Db 290 BERVDVSGEYANIRSMKVMFLFTDENWEISKLIAAILHGLNQLQYARFFENLDACEVIFS 349
Qy 351 TALGRSABLGLDLPQLTALTONSMPLRGSEILTLPLVANOAVSRDSRLAMALYACCEBW 410
Db 350 PSLATPAASILEVNPDLNMSCILTSRLITRGRTVSPSLREBALDVYDAFVVGIGYRLFW 409
Qy 411 VIKIKINSRI--KGNDFK-----SIGLIDIPGENEVNHPFQFNINYNANEXLOEYFNKHI 464
Db 410 IVDKINAAIYKPPSDVKNRSRISGLLDIIFGEFENAVNSFQOLCLINFANHEHQEFFVRHV 469

```



```

Qy 465 FPLEQLEYSREGIWMEDIMINGECGLDI-EKKGLLALINBESHFOATSTLIEKIH 523
Db 470 LTLDEEYDLESIDMIHIEFTDNOALMIANPMNIIISLIDBESKFPKGTDTMLHKIN 529
Qy 524 SOHANNHFFVKRVAANN-----FGVGVYAGEVQYDVGRGLIEKRNDFRDDLNLRESF 579
Db 530 SOHKLNANTIPPK-----NNHETQGINHFAGIYIYETQGFLEKRNDRDLHDDIIQLVHSSN 586
Qy 580 DEIYDLFEHVSRRNODTLKCSKRR--PTVSSQFVDSLHSLMATTSSNPFVRCIK 637
Db 587 KTKIKIF-----QADVAMGAEFTKRSPTLSSQPK-RSLBLMRTLGAQCPFFVRCIK 637
Qy 638 FVMQKMPDQDAVLINQLRYSGMLETVRIRKAGAVRRPFODFYKRYKVMRLA---L 694
Db 638 PNEFPKPYMLFDHMLCYRQLRYSGMMETIRIRRAGYPIRYSFVEFERYVLLPGVPAVK 697
Qy 695 PEDVYGKCTSLQLVYASNSMOLGKTKYFRESLEBQLEKREBEVSMAAVIRAHVIG 754
Db 698 QGDLRKTCORMABAVLGTDDMOIGTKIKFLKDHDMLLEVERDKAITDRIILLOKVIRG 757
Qy 755 FLARKQRYKLYCVVLIQ-----KNY----- 775
Db 756 FDRNSFLKLKNAATLIQHMNGHGNCRKRYGIMRLGSLRLQALHRSRKLHQOYRLARQRI 817
Qy 776 -----RAFLRRRFLHKKAAIYFQOKLREGIARVYRQLIAEK--REGEKKKQEE 826
Db 818 IQPQACRAVYLVRKAFRHLMAVLTVQAYARGMIAIRLHORLRAEYMLRLAEKQOLABE 877
Qy 827 EKKKE-----EEFERERERREALRKOGEETRKQOELALQSKQEAELTRELQ 880
Db 878 EKLIRKMSAKKAKEERBRHQERLQLABEDAE---RELKEKKAARKKELLEQMEBA 932
Qy 881 K-----ENKQVEIL-----RLEKEIEDLORMKQOELSTLEASLQLOQR 921
Db 933 RHEPVNHSMDVMKMFPLGTSGGLPGQESQAPSGFEDLRGR---MVEEDLDAALPL 988
Qy 922 RQDELRLLEE 931
Db 989 PDEEDDLSE 998

```

RESULT 9

AS9233  
 myosin VII-like protein - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C:Accession: AS9233  
 R:Anderson, J.B.; Yamashita, R.A.; Sellers, J.R.  
 submitted to GenBank, February 2000  
 A:Description: Complete cDNA for an unconventional Myosin (Class VII) in *Drosophila mel*  
 A:Reference number: AS9233  
 A:Accession: AS9233  
 A:Status: preliminary; not compared with conceptual translation  
 A:molecule type: mRNA  
 A:Residues: 1-2121 <AND>  
 A:Cross-references: GB:AF233269; PIDN:AAF34810.1  
 C:Genes: Myo28B1  
 A:Gene: Myo28B1  
 A:Cross-references: FlyBase:FBgn0040299  
 A:Map position: 2, 28B1  
 C:Superfamily: myosin motor domain homology  
 F:69-725/Domain: myosin motor domain homology <MMO>

Query Match 15.8%; Score 1697; DB 2; Length 2121;  
 Best Local Similarity 25.9%; Pred. No. 8.1e-69;  
 Matches 588; Conservative 362; Mismatches 847; Indels 484; Gaps 83;

```

Qy 5 GTRVLMRENGOHFPSTVNSCAE-----GIVFRTDYGQVFT---YKQSTITHQKY-TAM 54
Db 9 GEYVWVMPQ-----NTTSEPAVPFGARIVRTETQTLLVCDNRKQFWVVRAGDVLKAM 60
Qy 55 HTNNEGVUDMASLTLEHGSIMTNLFFQYKKNQITTYIGSIIASVNPQPIAGLYEPAT 114

```

```

Db 61 HITSQEDVEDMITTLGLOEYTLRLNLQNRVAKOLITYTYSMLVAINPYQ-ILPIYTRRE 119
Qy 115 MEQYRRHIGELPHIPIAIANECYRCLWKRRDNOCILIKGSSGAGKTESTKILKFLSVI 174
Db 120 IQLYKNSLAEIPHIPIAISDNAQRLORLKENQCVVSGSSGAGKTESTKILLOYLAI 179
Qy 175 SQQSLELSKERTSCVERAILLESSPIMEAFGNAKTVYNNSSRFQFVOLNICQKNIQ 234
Db 180 S-----KGHSIEWEQIIEANPIMEAFGNAKTVRDNSSRFQYIEIRPTGALQ 230
Qy 235 GRIVDCISSQNRVYRQGERNTHIFALLAGLEHERREBYL--STPENYHYNQSGC 292
Db 231 ARIQOYLL-EKSRIFYQSDERBNTHIFCYMLAGLSTARBRKQEOSSPYHYHIAQSGC 289
Qy 293 VEDKTIQSOFREYITMDVMOFSKEEYREVSRLAGILKGNIEFTTA-----GGAQV 347
Db 290 FTLPRGAKOPADIRAAKMLSPFEELVMSLSLAILHLGNIRFTATVANALATBEI 349
Qy 348 SFKTALGRSABLGLDPTQLTALTORSMPLRGEIITPLVVOQAVDSRDSLAMALVACC 407
Db 350 DDTPLQGVAAQLGIPISALNAALTORTIFVHGEHVTTLSKKAIEGRDAFVKSLYDGI 409
Qy 408 FEWYKTKINSRKMGED--FKSIGLIDFGFENPEVNHFEQPNITVYANEKQIYENKIF 465
Db 410 FVRIYRINETINKQVDQPMNSIGVLIFFGENPDNNSFEDLCINYANENLQPFVGHIF 469
Qy 466 SLEQLEYREGIWMEDIMINGECGLDI-EKKGLLALINBESHFOATSTLIEKHS 524
Db 470 KMEQDEYQENHINMOHIFQDNQOQLIDLIGKPMKMSLIEBSKFPKGTDTMLIEKHLV 529
Qy 525 QHANNHFFVKRVAANN--FGVGVYAGEVQYDVGRGLIEKRNDRFDDLLNLRESFDPY 583
Db 530 QHGNRSIYKGTQTSFGIRHAYGVVYMPGLFLEKGRDQSFQGLDITLVQSRSTNKYLV 589
Qy 584 DLFEEVSSRNODTLKCCSKRRPPTVSQFVDSLSHSLMATTSSNPFVRCIKRMQKM 643
Db 590 DIFPH---EMPMDTK-----KQPTLCVKFR-NSLDMRLRTLQAHPIYIRCIKENBYKE 640
Qy 644 PDQPDQAVVNLQRLSGMLETVRIRKAGVARRPQDPYKRYKVMRLMALPE--DVGK 701
Db 641 PGNFDELQVQLRISGMETARIRRAGYPIRHAIVAFVERLRLVPPVGLPQDCCKKL 700
Qy 702 CTSLLQLYDASNSEQQLGKTVFLRESLEQLEKREBEVSMAAVIRAHVYGLFARKQY 761
Db 701 ARQICEVALPADSDQYQKTLFLRDEBDABELEQSLMLKSIYTIQRGIRVLPFRYVM 760
Qy 762 RKVLVCVVIQKNTYAFILRRRFLHLKK-----AAIVFQOL----- 798
Db 761 KAYREALITVORVYMGRLQBRKRYQVMRQGFRLGACIAOQLTTKFTVWRCTTIQALIS 820
Qy 799 RGQIARRVYRQLIAEKREBEKKQEEKKKREBEER-----ERERERREALRKOGE 854
Db 821 RGYLVKDFQOKLERRKQNDQKEELIKLAKMEABELRLQLEKQERORQOEKR 880
Qy 885 TRKQOELBALQSKQEAELTRELKQENKQVEEILRLKEIIBDQRMKQOELSTLEAS 914
Db 881 LQEOQLKA-EAANANLAMAAVQOKRKTXYKQ-----EAPKAPTIQANSLPRPPT 933
Qy 915 L---OKLOERDDQELRLIE-----EBACRAOEFLEJLNPDEIDECVNIERSL 960
Db 934 LIVAAPLRTPRASATVTRINTIPESPTIDVSSKQWMDVFRFLN-DEPDALRLKLNNIS 992
Qy 961 SGGSEFSELEASACEKPNFNFQO-PYREEVDEGFADD--AFKQSPNPSHGHSQ 1017
Db 993 SGD---TRLRPSJPNNIDTSDFSYLKYAATYFGGATQHERKPKLKSLLKHEPIBEM 1049
Qy 1018 RTSGI-----RTSDSSEEDPYMNDTV-VPTSPASDSTVLLAFVQDSGLSNSS-SGS 1070
Db 1050 ASKAIWTLILFMGD-----LPDVSSPFLHVPFNENLMS-----DLASLNTSDSYKP 1098
Qy 1071 TYCMFQNAQDLPSPDGQYDYQDDYEDGATITSGSVTFSSYGSQMSPDYKSCVTYSS 1130
Db 1099 RLFFVQSORRIPKP-----LASGEKE--AOEFYQHM-----LNVFTSHLE 1136

```

QY 1131 GAYFSSSEG-AOSSPEDS--EEDFDSRFDTDELXYRSDSVSCVTLPYF-----HS 1179  
DB 1137 KIHFIIGHGIIKNSLRBEILAIQCKOYLNPSSRSYRGWLLSLCISCPREKEREPHL 1196  
QY 1180 FLYMKGG-----LMSWKRR 1194  
DB 1197 RSFMKGGTQLOATPSPLOLERTLVNGPCRPSPLELNAIRGRNPLKLDIHLMDCQORR 1256  
QY 1195 WCYLKDETFLMFRSKOEBALKO-----GWLHKKGGS-----ST 1227  
DB 1257 LOYDAAST-----AREAVOLOOSMGLTTFPGFGLVMSLNGKLMPLGAGOEIVLDAISE 1310  
QY 1228 LSRNNKKRW--FVLRSKLMYENDSEK-----LK-GTVEVTKAEIINDT 1272  
DB 1311 CEQRQLDAPKLYIRKEMFATWYDPSPMDPKATOLIVKQILNGKCEGYRCSRSEKDIAMVC 1370  
QY 1273 T-----KENGIDIMADRTPHLIESPED-----ASQFVLSQVHASTDDEIDQMD 1320  
DB 1371 ALACFVEYGGELIRLKPSEITFVPSDLAPGERAIEMNSRLIAATYKESVYKEQND 1430  
QY 1321 EQANPQNAVGTLDVGLIDSVCASDSPDRP-----NSFYIIPANRVLHCNADTPPEMHWT 1376  
DB 1431 LLEAQKRAK-----EDICLFAHLSWPRHSLRLEVV-----RKEGPKLQSDML 1475  
QY 1377 LLORSKGDTRVEGEEFVRGMLHKE-----VKNSPKMSLKLRWFVLTHNS----- 1424  
DB 1476 LGINSAGLFLIDETEOVLASCSEFVULKVHESDDKLVHVTFOHNVFVLQCSAQADANEV 1535  
QY 1425 ----LDYKSSSEKALKGLTVLNSL--CSVPPDEKIFETGYVNVVY--GRKCYRL 1476  
DB 1536 INYMLNLRQRSSYGAULDPVNGDLEDCLVLPNGDLEBAGVTAQALMAGNAQDCYR- 1594  
QY 1477 YTKLNEATWRSVIGVNTDTKAPITPQQLQIDIKENCLN----- 1518  
DB 1595 --GCVA--GOMGFLGANNRVALTLYKPSKQDILREGRFOEPKPYTPRANYSRRQHN 1650  
QY 1519 -SDVNEQIYK-----RNPILRYTHNPLHSPLLPYRGDINLNLKDKGYTLQDEAIKI 1571  
DB 1651 ISQLASHPREPLDSPKAPLSKPSPEPLKAPL-----KAVVAVY 1689  
QY 1572 FNSLQO-----LSMSD-----PIP---ITQGLTGNDLRPLRDELVQOLIKQTN 1614  
DB 1690 PPLFOQALVMMHHLIKYMGDIARSNLPVMTDLIFQPALQ--HPL--LCDELYQOLMKQJS 1745  
QY 1615 KVPHPSGVNLVSMQILITCSTFLPS-----RGILKYLKPH-----LKRIREQPGH 1662  
DB 1746 --DNPSSESSEKRMDDLILATGVAAPSVLVMEELIILMRADALADACLKRLKSLAOG 1803  
QY 1663 EMEKYA-LFTYESLKKTTC-----REFVPSRDEIALIHRQENTSTVYCHGGSCKITTN 1716  
DB 1804 ORKKAHLLIEVEGIQ--RCLHLYHKIYFPD-DTVEA-----FEIE 1841  
QY 1717 SHTTAEVVEKLRGLAMEDSRNMPALFEYNGHVDAISRTYVAVLAK-----FEKLA 1772  
DB 1842 SHTRGAEILADIAQRELK--SPVGYSIIFKTGDRVYVAMEEPEVFPIQLIYMLRQORT 1900  
QY 1773 TSEVGLPMKFYKLYCFGLDTPVPSKV--BPAFMEOAHNAVHGHHPAPEENQVLA 1830  
DB 1901 IRSISGQVQLMHMKLML--NNHPEGLNGDMIFSYPOELAKYLVGGYPIIDCEQSRIDA 1958  
QY 1831 ALRLQYLOGDYTLHAAIPPLEEYVSLQRLKARISOSTKTFTPCRLEKRTSLEGTLLR 1890  
DB 1959 IL-----VYSADHDVSLQRLP----- 1974  
QY 1891 SPTGSGVROKVEEOMLDMIMKEVSSARASIIDMKRFQGNQGNQAMAKWALIKEMP 1950  
DB 1975 ----EVLRLRLPE-----DLIPLOQVAERQOILPVRH--DHLTEHAAILFLOEISHTA 2024  
QY 1951 GYGSTLEFDV--ECKEGEPQOELMGVADAVSVYKRGEGRLVEYFOEHILSFGAPLANTY 2009  
DB 2025 CRGOSTFFVYKQONDALPETILAINSTGPHMLDPTTKELINSYEISQ--LGWISGSKNHF 2083

QY 2010 KI-----VDERELLEFESEVVDVAKMKRAYI 2036  
DB 2084 HIRFGNIGASKSLCTSTGYKMDLLASYV 2114

## RESULT 10

A59295  
unconventional myosin-15 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000  
C:Accession: A59295  
R:Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Ma  
an, T.B.; Frideell, R.A.  
Genomics 61, 243-258, 1999  
A:Title: Characterization of the human and mouse unconventional myosin XV genes responsi  
A:Reference number: A59295; MUID:20021762; PMID:10552926  
A:Accession: A59295  
A:status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-3511 <L1A>  
A:Cross-references: GB:AF144095; NID:96224684; PIDN:AAF05904.1; PID:96224685  
C:Genetics:  
A:Gene: MGI:Myo15  
A:Cross-references: MGI:1261811  
A:Map position: 11:33..9  
A:Superfamily: myosin motor domain homology  
F:1209-1871/Domain: myosin motor domain homology <MMO>

Query Match 14.8%; Score 1593; DB 2; Length 3511;  
Best Local Similarity 24.8%; Pred. No. 8,3e-64;

Matches 609; Conservative 373; Mismatches 836; Indels 642; Gaps 81;

QY 59 EGVDDMASLTELHGGSIMVNLFCYRKKNQITWTYIGSLIASVNPYQPIAGLYEPATMEQY 118  
DB 1205 EDGVDDMDQLBEDQOTTYLANLKTFRNLITYTIGSLIVSNPYRMRA--YGPQOVQY 1263  
QY 119 SRRHGLPEPHIFAIANECYRCMLWRHNOCLIKGSGAGKTESTKILKFLSVISQOS 178  
DB 1264 SGRALGENPHLFAIANLAFAMLDAKONQCVIIGSGSGKTEATKILRLAMNOR- 1322  
QY 179 LELSLKETSCYERAILTSSPIMEAFGNKATVYNNSSRFQGVOLNCOQNGCGRTV 238  
DB 1323 -----BDVMOQILBATPLLEAFGNKATVYNNSSRFQGVFIPL--EGVTCGALT 1373  
QY 239 DCILSSORVVRONGEENYHIFYALLAGLEHEEREYVLTSPENYHLYNOSGCVEDKTI 298  
DB 1374 SOYLBKSRIVQAKENNNYHIFYELLAGLPRLQRLQASLSQEAETVYLLNGNGNEINXG 1433  
QY 299 SPQESFREYITAMVMOFSKEBEVRSRLAGILHGNIEF-----ITAGGAQVSFKTAL 353  
DB 1434 SDADDFRRLAMEVLTGFTSEDDOSIFRILASILHLGNVYFEKHETDAQEVASVVSAREI 1493  
QY 354 GRSABLLGLDPTQLDALTORSMPLRGEIILTPLVQAQAVDSRSLAMALVACPEWYIK 413  
DB 1494 QAVALLQVSPGLOKALATFKVETIRKIFPLPLVESAVDARDIAVLVALLFGMLIT 1553  
QY 414 KINSRIKNEDEPKSIGLIDIFGFENFEVNHQFINANEXLOEYFMKHIFSLQLEYS 473  
DB 1554 RNALVSRKQDPLSLAIDIDYGFEDLSFNSFQOLCINANENLOVLFRKIYVQEOBEYI 1613  
QY 474 REGLVWEDIWDINCECLDLB--KKGLALLINEBSHPQATSDTLEKLHSQHANHFY 532  
DB 1614 REQOMREIAPADNPQCNILSLKPYGILRLDDCCFPQATDHTFFLOKCHYHGANPLY 1673  
QY 533 VKPRVAVNPFQVKYAGVOYDVRLGLEKNNDTPRDDLLNLBRSRPFYIDLFHVSER 592  
DB 1674 SKPKMPLPEFTIKHAYGATVQVHKFELDKNDQVQVDLDFVHSRTVVAHLFSSHAQ 1733  
QY 593 NNQDTLKCGSK-----HRRPTVSSQPKVDSLHSLMATLSSSPFRCIKPMNQKPPDOF 648  
DB 1734 TAPPLGSSSITRLRYKATYAAKQ--OSLDDVVKMRCHNPLFRCLKRNHKEPGLE 1792  
QY 649 QAVVLNQRLYSQMLTVAIRKAGAVRRPFDQFYKRYVLMK--NLALPBDVARGKCTSL- 706

```

Db      1733 PDVWMAQLRYSGLVTLVRIRKEGFVRLPFQVIFIDRYRCLVALKLVNPRD--GDMCVSLIS 1851
Qy      707 QLYDASNSBMOAGTKTVFLRESLLEOKLEKREESVSHAMVRAVHLGFLAKOYRKLY 766
Db      1852 RLCTTTPDMYRGVIGISKFLKEHLHOLESMERVOQRALVTLQRLIRGFTRHRSLSR 1911
Qy      767 CVV11QKNYARFLRRRFLHLKKAIVFOKOLRGQIARRVRYQLAEKKEQ--EEKKOE 825
Db      1912 KIILQSRARGFILAQORVQOMOSLLKFRSLVIVNRRRYLKLAEOGRRAQEMLRQ 1971
Qy      826 EEKKKREBERERREBRRAELRAQOEETKQJELALQSKAEELT--RELEKQKE 882
Db      1972 BELSKREVV-----PVRLBVPA-----EVAQGLLOAAAGLKSSGGRVAVVPA 2015
Qy      883 NKQVEILRLKEIEDLQWKEQELSLTEASLQKLEGERDELRLLE--EACRAQEL 941
Db      2016 RLQAEPCVTLPLDINNYP--MAKFRCHFEKPEPFCHLTVPLKMPRLRPEVHNAELISVFK 2074
Qy      942 ESLNDEIDEVCNTERLSGSEBESSLEASACEEK-----PN--FNFSQPY- 987
Db      2075 LILRFMG--DPRHLGQEMILGNVYHOGVLEPALRDEILAQLANQVWRPNAYNSRGL 2133
Qy      988 -----PEEVEDE-----GEADDAFAKDSNPSEKHSDOGTGIRTS 1025
Db      2134 LLAACLGPAPSPHLDKFLKEVDYQNGQFQA-----VCOHRLLOAMGSGAART 2183
Qy      1026 DDSSEEDPYMNDTVPTSPASDSTVLVLPASVODSGLSHNSGSESTYCMQPNAGLPSPD 1085
Db      2184 PPTQ---LEWTAIOEKASMLDVSCEFNGDQFSCPVHTWSTGEAV-----AGDLKIR 2233
Qy      1086 GDYDODDYEDGATITSGSVTFNSYSGSQSP-----DYRCSV-----GTYN 1128
Db      2234 G-----LADG--WRGWTAMKN--GYQWABELIGHDVLVDLVDLDELRLRPPKQSKF 2281
Qy      1129 SSGAYRFSEGAQSSFEDEEDPDRPTDELSTRDSVYCVTLPTFHSFLYMK--GG 1186
Db      2282 IVGA-----EGPLAGDTRGVFGNCGWDSDEDTPTRPQPDHVAKMPDLGCSHEDST 2336
Qy      1187 LMSMKRW-----CYLKDEFLMFRSQ 1210
Db      2337 NGETAQRTSNRQAVDSIGESTVPRBLDGYLSLFDVLAAGDALKKPTALIAIRMK- 2395
Qy      1211 BALKQWMLHKGGSGST--LSRRWK-----KMFVLROSKLWYEND 1251
Db      2396 ---GGGPGGGGGSTSEDTSRPRPEKLPKPIPGIDASTLALQAFIRQAVLLAREMT 2450
Qy      1252 -----SEKLGKT-----VEVR-----TAKSIIDNTYEN----- 1276
Db      2451 LQALALQOQPLSATSRPQLPRERPLAERAPKTVGVGPAPKAPVLVPRTPQSWAPGSVAKA 2510
Qy      1277 -----GIDIMADRTF--HLIASBPDAQWFLSVLQVH--ASTDQELQEM----- 1318
Db      2511 PKIPKPAVAPILADQWTPAESIASPELVR--YSTLSEHNPPOPTQOIRSIIOKQORP 2568
Qy      1319 -----HDE-----QANPQNAVGTLDVGLDSV----- 1340
Db      2569 WAGHREARTDGGKVRPRPDHEEALMLKQKQQLAVPFGVSRVAVNAVKPTSPAR 2628
Qy      1341 --CASDSPRNSFV-----ITTAN----- 1358
Db      2629 RQCMGPTVPQPSRLEPRDPVOTQLRLVNPFGYQDIPIRIFLKEVFPKQNVSHP 2688
Qy      1359 -----RVLH-----CNADTPREMHMITLLQSKSDTVVEGEFVPRGMKEVKN 1404
Db      2689 VOLDLFRQILDHTPSEACLRISEDERLOKALFAQNDLDTQ-----RLVYVESVXR 2740
Qy      1405 S-----PKMSSL-----KLKRWFLV----- 1420
Db      2741 AAIISARDSWEIYFSRLFPAMSGVGTGVDILAVSHTGICKLQWVKSASRLRLVLCY 2800
Qy      1421 -----THNSLDYKSEK-----NAIKGLTIV-----LNSLCVV 1450

```

```

Db      2801 SPADILEVTMPBSQNMLEFNLSNEKLIIPSAQAQVKTLVTFTILKKDSYVAVARNFL 2860
Qy      1451 PPEKIFK-----ETGYWNVTVYGRKCY----- 1474
Db      2861 SEDEBLSFHKGDIIHLQSLPTRVGVSAGCVAKKLVLEELRRRGDPFGMRGAVHGR 2920
Qy      1475 --RLYTKLINBAT-----RWSVIONVDTKAPIDT-----PTQQLIOD 1511
Db      2921 VGRPFSELVQAAABDFQLPAEPRGRAAVVAVAASAAQAGRRREGPRVYARASAD 2980
Qy      1512 IKENCL---NSDVE--QIYKRN-----ILRTYHHPHSPL 1543
Db      2991 SGEISALPSTMEFAKQYPRDRPRRDLKLSKEDRSKLTLEDVLCFTKVPIDQSL 3040
Qy      1544 LPLPYGDINLMLDKGYTTLQDEAIKIFNSLQO-----LESMSDPIPIIGILQTSHD 1597
Db      3041 IEL--SDSNLNM-----AVDMFVAVMFGMDAPLKQGS--LDVLTCLLKCGD 3086
Qy      1598 LRLRLDELQCLIKQTNKVPHPGSGVNLVSMQILTCLSCFLPARGILKYLKPKLRIE 1657
Db      3087 HEVNRDECYCQIVKQITDSSPKDSCQGRWRLIYIMAAVYSCSEVFYPYLIRFLOHV-S 3145
Qy      1658 QPPTMEKVALFTYESLKT---KCRFVPSRDEIELI--HQEMSTVYCHGSGSKI 1713
Db      3146 WTPGLPFQGIKACEQNLQKTLRFGRLEPFSNMELRLMLAGRSKQSLPGLEHRL 3205
Qy      1714 TINSHTAGEVVEKLRGLAME-----DSRMWFLFENGVDKALIESRTVADVLAKE 1768
Db      3206 KIKCTVALDVIEGLCTEMALTREBAPDEYIIFVYTRNGQV--CPLSRAVYILDVASME 3264
Qy      1769 K-----LAATSEVDDLFWKFFKLYCFLDTDVPKQSVBEAFMEQAHBAVTHGNHAP 1822
Db      3265 QVDGCVTLMPFRVLMVDQGLKRENELVYTMHNOVLPLYLKGIF-----SSVAPRGTE 3317
Qy      1823 EENQVLAALQYIQGYTTHAIPLEBYYSIQRLKARISQSKTKTTPCEBLEKRTS 1882
Db      3318 QOLQOVSKLALQHRAXD---HFLVPSVREVO--EYPAQVYHTAG----- 3359
Qy      1883 FLEGLRSPRTGVSVVRQKVEEQGLDMWIKEEVSSARASIIIDKWKFGQGNQOAMAKY 1942
Db      3360 -----DTWL--NLVSOHR-----QOTQALSPPHQAQOF 3385
Qy      1943 MALIKEMPGYSTLEPVE--CKEGFPQELMGLVSDAVS-----VYKR 1984
Db      3386 LGLSAPFLPGSSPFPIQSGNVLVPAFCILAVHNGINPLSTKTHLIVKIPKEIGST 3445
Qy      1985 GSGRPLEVQYEH--LSFGAPLANTYKIVDERELFETSSEVDVAKMKAMKAYISIVYKR 2043
Db      3446 WTKETANSSYPVVEISIGD-----VAAORTWQLOEBGLCRVAVAVHVESMLSAR 3497

RESULT 11
A59266
unconventional myosin-15 - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59266
R:Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Mi-
an, T.B.; Fridell, R.A.
Genomics 61, 243-258, 1999
A:Title: Characterization of the human and mouse unconventional myosin XV genes responsi-
A:Reference number: A59266; MUID:20021762; PMID:10552926
A:Accession: A59266
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-3530 <L1A>
A:Cross-references: GB:AF144094, NID:G6224682, PIDN:AAFO5903.1, PID:G6224683
C:Superfamily: myosin motor domain homology
F:1225-1887/Domain: myosin motor domain homology <M0>

Query Match 14.8%; Score 1589.5; DB 2; Length 3530;
Best Local Similarity 25.1%; Pred. No. 1.2e-63;
Matches 623; Conservative 348; Mismatches 818; Indels 697; Gaps 88;

```

QY 59 EECVDDMASTLTELHSGSIMYNLFORYKRNQIMTYIGSILASVNPYOPIALGLYEPATMEQY 118  
 Db 1221 EDGEVDMTOLEDOETVLSNLKIFRERNLITYIGSILVSNPYO-MFGIYGEVQVOQY 1279  
 QY 119 SRHGLDELPHIATANECEYRCMLKXHDNOCILIKESGAGKTESKTLIKELSVISQOS 178  
 Db 1280 NGRALDENPHILPAVANLAFKMLDAKONOCIIISESGSGKTEATKLILRYLAANNOK- 1338  
 QY 179 LELSLKETSCEVRALESSEPIMEAFQNAKTYVNNNSRFGKFEVONICOKNIGOGARV 238  
 Db 1339 -----REWMOIKR-ILEATPLLESFGNAKTYVNNNSRFGKFEVIFL-EGGVISGALT 1389  
 QY 239 DCILSSQNRVVRONPGEERNYHIFYALLAGLEHEEREFEYLPSTPENHYLNOGSCVEDKTI 298  
 Db 1390 SYLLEKSRIVFOAKNERNYHIFYELLAGLPAQLROAFSLQBAETYYILNOGNCISIAQ 1449  
 QY 299 SDOESFREVTIANDWQFKEEYREVSRLAGILHGNIEF-----ITAGAOVSFKTAL 353  
 Db 1450 SDADDFRRRLAAMEVIGFSSSEODSIIFRLASILHLGNVFEKXETDAOEVASVAGARI 1509  
 QY 354 GRSAGELGLDPTOLTALTOBSMFLGEEILTPANTQOAVDSRSLAMLYACCFEMVIX 413  
 Db 1510 QANAEILOISPEGLQKAITKTETREKIFTPUTYESAVDADALAKULYALLFESWILT 1569  
 QY 414 KINSRIKGNEDFKSIGILDIFFGEFENFEVNHFEQFNINYANEKLOEYFNKHTFSLQLEYS 473  
 Db 1570 RVALVSPRODTLSIILDIYGFEDLSFNSFEQLCINYANENLOYFNKIVFQEEQOEYI 1629  
 QY 474 REGUWEDIDWINGECLODIE-KUGLALINEBHFPOATSTLTLEKLSGHANNHY 532  
 Db 1630 REGIDMOETEPANQPRINILSLKPYGILILDDOCCFPQATHTLQCKHYHGNPNLY 1689  
 QY 533 VKBRVANNNEGVAGVEVOYDVRGLEKKRDTFRDLDNLLEBSRFDYDLF-EHVS 591  
 Db 1690 SKRKMPLPETTIGYVAGKTYOVHKEFLDKXHDQVROVDLFPKSTRVVAHLFSSHQ 1749  
 QY 592 RNNQDTLKCSG---KHRRPTVSSQFYVDSLHSLMALTSSNPFFVACIKRNNQKMDQDF 648  
 Db 1750 AAFORLGKSSVTRLYKAHTVAAKFO-QSLDLVEKMERCNPLFMCLKNHKEGFLPE 1808  
 QY 649 QAVVNLQALRYSGMLETVRIKAGYAVRRPRODYYKTYKULMRLLAPEDVGRK---CTSL 705  
 Db 1809 PDVVAQALRYSGVLETVRIKKEGFPVLRPEQGFIDYCCCL---VALKHLRANGDMCVSV 1865  
 QY 706 L-QLYDASNEWOLGKTYVFLRESLECKEKRESEESHAAVIRAHVIGFLARKOYK 764  
 Db 1866 LSRLCVMRMYVGVSKFLKCHLYOLESKREHYNLALTLQCLGFTIKRFRSL 1925  
 QY 765 LYCVVILQKNYRAFLRRFLKKAIVFOKQLOIARVYRDLAEKREO-EKKKQ 823  
 Db 1926 RHKIILLOSARGYLARQYQWRSLVFRSLVHAIVSRRLYKLRAEWRQVEGALLM 1985  
 QY 824 EEEKKKRE-----EEERERERERERELRAOQEBET 855  
 Db 1986 EOBELSKREVAVAGLEVPRAELAGLQAVAGLGAOVAPVPTBRLOAEPRTVLPLDI 2045  
 QY 856 RKQOELALQKSQK---AELTRELE-----KOKENKQVEEILRLEKEIEDORMKEOQ 906  
 Db 2046 NNYPMAKFYQCHKEKAPFGMLTYPLRPLQOLRAHHAERAVSIFKILAFMGDPHILHGR 2105  
 QY 907 ELSLTSASLOK---LOERRDOELRLEEBACRAAOEFLSFLNDEIDECVRNIERS--- 959  
 Db 2106 ENIFGVNIVQKGLAVBELDEILAOQANQVMH-----NNAHAHAERGMILL 2151  
 QY 960 ----LSGGSERSSSLASACEKRNPFQSPREEVEDEGEADDDAFKOSPNSEH--- 1012  
 Db 2152 AACLSG-----FAPSFCFNKTYLKFVSDYGR---NGFOA-----VCOHRLMQ 2190  
 QY 1013 --GHSQORTSGIRTSDSSEEDPYMNDTVVTPSPADSTVLLAPSVQDSG-----S 1061  
 Db 2191 ANGRADQOQSSGAAR-----TLPTOLENTATYKASMLADVCGCRNGDOFSCP 2237

QY 1062 LHNSSGSESTYCMQNAQDLSPDG-----DYDQDDYED-- 1097  
 Db 2238 VHSWSTGEEV-----AGDILHHRGLADQMRGVTAMKNGYQWABELAGHDVLDLVSDLE 2291  
 QY 1098 -----GATSSSVTFNSYSSQSPDYRCGV----- 1124  
 Db 2292 LIRDPPOKSYITVETEGPALSRRGPKVFGNS-----WSDSDMSTRPOOEHRPKVLDS 2347  
 QY 1125 ---GTYNSSGARYFSSEGNOSFFEDSED-----FDSRFDTDDELSTYRD 1166  
 Db 2348 DQSYSHNDQGT-NGETEAQRGTATHQESDLSGEPAVPHKGLDCYDLSLF--DPVLSYDGA 2404  
 QY 1167 SVYSCVTLPEYHSLYMKGLM-----NSWRKRCVLK---DETFLWRSKQBA 1212  
 Db 2405 DLEKPTALAY-----RMKGGCQPGCGSSGTEDTFRPRPEPKPIBGLASTL-----A 2452  
 QY 1213 LKQGLHKKGGSSSTLSRRMKKRFVLRQSL-----MYFENSEEKLKGTVEVRT-- 1264  
 Db 2453 LQOAFTHQO---AVLLARGMTLQATLQOQPLSALSLPAEKPPABEAOPT-SVGTGP 2507  
 QY 1265 -AKEIIDNTT-----XENGIDIIMADRTFHLIASEPDAEQ-----WF 1301  
 Db 2508 PAKPVILKATPKPLAPAPLAKAPRLPIKPVAPVLAQQA-----SPETTSPELVRV 2561  
 QY 1302 SVLSOVH-ASTDOEIOEM-----HDE----- 1321  
 Db 2562 STYNSEHPPOPTQIKNTIVROYQOPFRGGRPEALRKDGKVMKRPDHEALMILKQOM 2621  
 QY 1322 ---QANPONAVGTLVDGLDSVCAQSDP-----RPN 1351  
 Db 2622 THLAAPGTQVSREAVALVKPTSAPRPSMAPTSLPBRSLPEBELQOTRLHRLINPVF 2681  
 QY 1352 F-----VIITANVLH-----C----- 1363  
 Db 2682 YGOQAPKPIFLRKEVFPKOSYSHPVQDLDFRLIHLDTLSEALRISDEBRLMKALF 2741  
 QY 1364 ---NADTPBEM-----HMITLLQ 1379  
 Db 2742 AQONQDLOKPLVTESVKRAVSTARDTWEVFSRIFPATGSGVGVQLLAWSHVGIKLLR 2801  
 QY 1380 RSKGDTVRGGEFPIRGMLKEVKNKSPKMSLSLKKKRFVL--THNSLDYKSSBK----- 1433  
 Db 2802 MYKGGQEBAGQOLRVRAVSPADI-----LVTPSQMULEFNLASEKVLIF 2847  
 QY 1434 --NALKGLTV-----LNSLSVVPDEKIF-----KETGY- 1462  
 Db 2848 SARBAQVTVIYDDFLELKQSDVYAVARNFLPDBPALAFHKGDIILHQLPEPRVGS 2907  
 QY 1463 -----MNV-TVGRKHCYRLYTLNLNEAT----- 1485  
 Db 2908 AGCVRRKVVYLEBLRRRGPDPFGMRFGTIGHR--VGRPSEIVQPAAPDFQLPTEBGR 2965  
 QY 1486 -RMSVIGNVDTKAPIDT-----PTQOLIODIKENCL--NSDYVE--QIYKRN- 1530  
 Db 2966 GRAAAVALAAVSAALAOEVRRRREGPVRARSADGBEALALPYTLMLEPAKTRIDROR 3025  
 QY 1531 -----ILRTHNPLHSPLPLPYGDIINLNLKDKGYTTLQDBAI 1569  
 Db 3026 RPQDGLRUKSKEPRRESRTLEMLCTKTPLOESLEL--SDSSLSKATDWFL-----AV 3078  
 QY 1570 KIFNSLQOESMSDPIPIIOGILQTHDRLRPLRDELYQOLIKQNKVPHPSGVNLYSWQ 1629  
 Db 3079 MRFMDAPLKGQSD-LDVLGNLKLKCGHEWVRDECYQOVVQOITDNTSSKODSCORQMR 3137  
 QY 1630 ILTCL-----SCTPLPSRGILKYLKFHLKRIE--OPFGTEMBKALFTYESLKKT--K 1679  
 Db 3138 LUYITAVHSCS-----EVLHPLTRPLQDVSRTPGRLPPOGIAKCBONLQTLTLPFG 3189  
 QY 1680 CREPEYRDEIEALI-HROEWTSTVYCHGGGSCKITINSHTTAGEVVEKLIRGLAM--ED 1736  
 Db 3190 GRLEPSSIELRAMLAGSSSKROQLFLPGLGEBRHUKITCTVADLVBEICAMALTLTPRE 3249  
 QY 1737 SRNMFALFEYNG--HYDKAIESRTVADVLAKEEK-----LAATSEVGLDPMKFFRYL 1787



QY 6 TRVLRENGOHFPSTVNSCAE-----GIVFRDYGVFTYKOSTITHOKVTAMHP 56  
 Db 11 TRWIPD-----PEVVRSAHELTKDYKDGESDLOLRJEDDTIIDYPRIDVQNVQVPLRNP 65  
 QY 57 TNEBGVDDMASLETLHGGSIMVNL-FORY-KRNOIWTYIGSILASVNDYQPIAGLYEPATM 115  
 Db 66 DILVGEENDLTALSHLHPAVLHNLKVFLESNMHIYTCGIVLVAINVEQLP-IYGDQVI 124  
 QY 116 EOYSRHLGELPHPIFALANECYRCLMKRDNOCILKGGSGAGKTESTLILKFLSVIS 175  
 Db 125 YASGQNMGMDDPHITPAVABAYQOMARDEKNOISIVSGESGAKTVSAVYARMEYFTVG 184  
 QY 176 QOSLELSKEKTSCEVERALIESSPIMEAFGNATVNNSSRFGEVOLNICOKNIOGG 235  
 Db 185 GSAADNTIEEK-----VLASSPIMEAIGNAKTRINDSSRFCKYIEIETFDKXHYHIGA 237  
 QY 236 RIYDCILSSONRVVRQNGERNYHIFYALLAGLEHEBEREPYLTSPENHYLNQSGVED 295  
 Db 238 NMRTYLL-EKSRVVFQADDERNNYHIFYOLCAASLPEFKELALTCAEDFYTAHGGNTTI 296  
 QY 296 KTSIDOSFEVITAMVMOFSKEEVREVSRLLAGILHIGNIEF-ITAGAQVUSFKTA-- 352  
 Db 297 EGVDDADPEKTRQALTLTGVRSHQISIFKTIASILHLSVEIQERDDSCSISPODE 356  
 QY 353 -LGRSAILGLDPTOLTALTORSMFLRGEEILTPLVQOAVDSRDSIALMALYACCFEW 411  
 Db 357 HLSNFCLLGIBHSQMHMLCHRKLVTTSETYVTKMSLOQVVARANALAGHYAOLFSM 416  
 QY 412 IKKINSIKGN-EDFSGIGLIDFGFENFEVNHQGNINIVANEKLOEYVNHKIFSLLEOL 470  
 Db 417 VERINKLQTSLOKSHSPFGLVDIYGFEIINSFEOFCINIVANEKLOOQFNSHVFLEOE 476  
 QY 471 EYREGLVWEDIMIDGECLELIEKGLGLALINESHPPOATDSTLLEKLSOHANN 530  
 Db 477 EYKGOIPLTIDPYDQPCIDLEAKGLIDLEDECKVPKGTDOMAOKLERSNSO 536  
 QY 531 FYVPRVAVNPGVKHAYGEVOYDVRGILEKNRDTFADLLNLRESRPFIDFL--- 586  
 Db 537 HFQKPRNSNTAFVIHPRADKVEYLSDFLEKNRDYEEQINILKASKPFLVADLRDDE 596  
 QY 587 -----EHVSSRNNODTLKCGSKHRRPVTYSSQFKVDSLHSLMATSSSNPFF 632  
 Db 597 DSVPATNTAKSRSSSKINVRSSRPLMKAPNKEHKSQYQFRT-SINLLETINATTPHY 655  
 QY 633 VRCIKPMQKMPQPODAVVLNQLRYSGLJETVIRKAGYAVRPPODPFKRYKVL--R 690  
 Db 656 VRCIKPNDKLPHPFDKRAVQOQRACGVLETTIRISAGYPSRWYTHDFPNRIRVLMKKR 715  
 QY 691 NLALPEVDRGCTSLQLYDASNEWOLGKTVFLRESLEOKLEKREBEREESHAAVIRA 750  
 Db 716 ELANTTDKKNICGSVLESILKDPDKQPGRTKIFPAGQVAYLEKLRADKFRATIMICK 775  
 QY 751 HVLGFLARKOYKRVLCVVIYQONRAFLRRRFLHK--KAAIVQOKOLRGCIARVY- 807  
 Db 776 TVAGWLQRYKVRRLRAITLQRCRGYLAIRLLEHLRRTRAIIVQOKOYRMLKARRAVC 835  
 QY 808 ----- 807  
 Db 836 RVRAAVIIOSTYRGHVCTOKLPPVLTCHKATIIQYKANGMARRHFORORDAIVIOQA 895  
 QY 808 -ROLLAEKREOEKKOQEBEKKKREBEREREREREAELRAQOEETRAKOELBALOK 866  
 Db 896 FRLLKARQALKALKEARSSEHLKRLNVGMENKVUOLQRKIDQNKNEPFLTSQLSAIVS 955  
 QY 867 --SOKAEILTRELKOKENKQVEIIRLEKEIEDLOPMKEQOELSTIASLOKLOERRDO 924  
 Db 956 THAMEVERKKEKELARQONOEADPSLOLEBEVQSLR--TELQVMSERVLJEHANRRENG 1013  
 QY 925 ELUR-----LEEBACRAOEFLBSINDEIDECVRNTERLSGGSSESSSELSAAGECKEN 980  
 Db 1014 ELAKRAVADLEHNAALKDE-KEHLN---HQIURKSSASSOSSVEENLILKKELEERS 1068

QY 981 --FNFSQPYE-BEVEDGEFDDDAFKOSPNESEHG--HSDORTSGIRTSDDSSSEDPYM 1035  
 Db 1069 RYONLVKEYSOLQORENENRDSQOTPPGHRKNRNSGSLSDSNYSISITSEIGTDEDAQ 1128  
 QY 1036 NDTVPTSPSA-DSTYLL-----APSVODSSSLH-----NSSGSESTYCPMONGDLPS 1083  
 Db 1129 QVEEIEIEKAAMDVTFELQKRVLEBEQRKLDVQLEKEQDSKQVQVEOQNGLDVD 1188  
 QY 1084 PGDDVDQDDVEDGATSGSSVTFSSNSYGQMSPDYRCSVGTYNSSGAYRPSSEGAQS 1143  
 Db 1189 QDADIAYNSLKQOE--LESENKKLNDLNEWK-----AVADQAMQDN 1229  
 QY 1144 FEDSEDFDSRF-----DTDELSTYRDSVSVSVTLPYFHSFLYMKGLMNSMKRVCVL 1198  
 Db 1230 STHSSPDSYSTLLNQKLKANELLEVREKVEV-----LILRTQIMANDQRLSLGK 1277  
 QY 1199 KQETFL-----WFRSQEALKQGMHLKKGSGSSTLSR-RNKKRRFVLRQSLKMTFENDS 1252  
 Db 1278 NMEPNINARTSWPNEKHHVDODALEAHYGVQVTSQTEDMG-----YLNBDG 1325  
 QY 1253 E-----EKLGVETRTAKEIINDNTTKENGIDIMADRPHLJASPEDASQMF----- 1301  
 Db 1326 ELGLAYQGLKQVARYLLEAOLQONLKHREVEHLQAO--VEAMKEBMDQOQTFQOTLL 1383  
 QY 1302 SVLSQVHASTDOEIOEMHDEQANPONAVGTLD-----VGLIDSVCASDS 1345  
 Db 1384 SPEAOVEFGVQOEIRLTJENENDPFKEVLEKJENKELKKOLKIYMKVQVDEAQAOLAO 1443  
 QY 1346 PRPNFSVLTITANRVLHCNADTPREEMHWT--LLQSKGDTREVGQEFYRGUL--HK 1400  
 Db 1444 SRR-----RHHELTQVTVORERKD-----FOGMLEYHK 1472  
 QY 1401 E-----VKR-----SPKSS-----LKLKRFVLTNHSLDYKSEKXALGLTVLNS 1445  
 Db 1473 EDEBALLIRLUTDLKPOLMSGTVRCLPAYILMYCIRHAD--YTMDLKVHSLSTING 1529  
 QY 1446 LCSVPRPEKIPKETGVNNTVYGRKCYRLYTKLLNEATRWSSVYQNTDTKADIDPT 1505  
 Db 1530 IKKVLKKNHEDPEMTSFMLSNTRCLHCLKQYSGDEGFWTQ-NTAKON----- 1576  
 QY 1506 QOLIODIKENC-LNDDVBOIYKRNPIRLYTHNPLHSLPLRPYGDINLNLKDKGYTTL 1564  
 Db 1577 -----EHCKNNDLFE--YRQ-----VL 1592  
 QY 1565 ODEAIKIFNSLOOL-ESMSDPIPI-----IOGILQTHDLRP----- 1600  
 Db 1593 SDSLQIYQOLIKIABGLQPMIVASMLNESIQLSGVRPFGYKSSSWVDGNSYCL 1652  
 QY 1601 -----LRDELVCOLIQTNNKVPHPGSVGNLYSMOILTCLSTFLPSR 1642  
 Db 1653 EAIIRQMPFHTVLCDOGLDPEIILQVFKQLPYMINAVTLNLL-----LLRKDACSM--ST 1706  
 QY 1643 GILKVLKHLKRIKQPFQTEMEKY-ALPTYESL-----KTKCREVPSRDELE 1691  
 Db 1707 GM--OLRANISQLEBMLRGKMLQOSGAVQTEPRLQAAQLOLKKT-----QEDAE 1756  
 QY 1692 AL-----IHRQEMSTYVCH--GGSGCKITIN-SHTTAGEVNEKILRGLAMEDSRNMF 1741  
 Db 1757 AICSLCTSLSTQOIKILINLYTPPLANGFEREVTVSIRITIOAOLQDRSPDOLLDSKMF 1816  
 QY 1742 -ALFEYN 1747  
 Db 1817 PVLFPFN 1823

RESULT 14  
 A26655  
 myosin heavy chain [bimilarity] - alime mold (Dictyostelium discoideum)  
 N:Contants: myosin ATPase (EC 3.6.4.1)  
 C:Species: Dictyostelium discoideum  
 C>Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 19-Apr-2002  
 C:Accession: A26655, A24728, S00250  
 R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.





QY 70 ELHGSIMYNL.FORYKXNOIWTYIGSILASVNPYOPITAGLYEPATMEQYSRHGLBPH 129  
 Db 156 ELFEENLKLKHLKFLKOOKIYTAGSILVAINPFR-PLPIYKPYKVMYENQOLGKLEPH 214  
 QY 130 IFPAIANCYKMLKRNOCILIKGESGAKTBTSTKILKFLVTSIQOSLELSKXTSC 189  
 Db 215 VFLADAVAYTMYLKRKNOCIVYPGESGKQSTYNLHCLTALSG-----KGYASG 267  
 QY 190 VERALLESSEPIMEAFNAKTIVNNNSRFGVOLANICQKNIQCGSIVDCILSSQNRV 249  
 Db 268 VERITILGACVLEHFGAKTAHNNNSRFGKFIQVSTLBSGYRGA-VAKYILLEKRLV 326  
 QY 250 RQNGERNYHIFVALLAGLEHEEREFPYLPENYHYNOSGVEDTISDQ-----SFR 305  
 Db 327 SOEDERNHYFVYLLGVSEERQEFOLKOPEDYFLLQ-----HNLKIDEGDLKHDFE 382  
 QY 306 EVITAMVMPQSKBEVNEVRLAGILHNLNIEITNG-----GAQVSFTALGRSHELL 360  
 Db 383 RLKQAMEMVGFPLPATKQIPAVLSAILYGNVYKRRATGREBLEGEPREVLDTLSQL 442  
 QY 361 GLDPTOLDALTORSMFLRGEELITPLANOAVDSRSLAMALYACCFEVIKINSRIK 420  
 Db 443 KVRKEILVEVLTAKKYTVNDKILPYSLEAITARDMAKSLYSALFDMIVLRINHAL 502  
 QY 421 GNEDFK-----SIGLIDIFGFENFEVNEFOFNINANEKLOEYFNKAI FSLDELYSR 474  
 Db 503 NKLDVEBAVCSIGVLDIFGFEDFERNSFEQICINANBOLQYFNOHIFKLEOEYOG 562  
 QY 475 EGLVMEIDWIDNGECIDLIEK-LGLLALINESHPQATDSTLEKHSQANNHYV 533  
 Db 563 EGTWNIHIGYTDVNGCIHLISKKPTGLFYLLDESNPHATSOQLLAKFQOQEDNKRYFL 622  
 QY 534 KPRVAVNPFVCYAGEVOYDVRGILEKNRDTFRDDLNLRESRFDYIDL-----586  
 Db 623 GTVMEBAFIIQHPRAGVKYQIKDFREKNMDYMRPDLVALLRGSDDSYABELGMPVAV 682  
 QY 587 -----EH-----588  
 Db 683 FRMAVLRAIRMAVLRERGLRBAERAKAGMSSPQAGSHPELPRGASTPSEKLYRDL 742  
 QY 589 -----588  
 Db 743 HNOIKISIKGLPMOGEDPRSLSLOSLRQPRAFILKSKGIKOKIIPKXNLSKSKL 802  
 QY 589 -----VSSRRNODTLKCGSKHRRPTVSSQFYDLSHLMATLSSSNPFVVCIRKPNQKMP 644  
 Db 803 ISMTLDRRTKSLHLHKKKKPSISNQFOT-SLKLLEALGKAEPFRICISMAKEKE 861  
 QY 645 DQFDQAVNLQRLYSGMLETVRIRKAGYAVRRPDPYKRYKYLNRNLALPEV---RGK 701  
 Db 862 LCFDDELIVLOQLRYTGMLETVRIRRSQYSAKYTFQDFTBOQVY-----LPKDAQCREV 916  
 QY 702 CSTLLQLYDASNEBMOGLTKVFLRBSLEKLEKREBEVSHAAMTYIRAHVGLFARKQY 761  
 Db 917 ISLLEBMTIKDKNYOIGKTKVFLKETERQALQETLHREVVARKILLQSGFRMVLRRHF 976  
 QY 762 RKVLCVVIIOKNYRAFLRRRFLHLKKAIVFOKOLRGQIARVYR-----OLLA 812  
 Db 977 LQKRAAVTIQACMRGRVRA-LERTQAAVYLQAAWGYWOKLIRHQKOSTIRLOJSLC 1035  
 QY 813 EKREOEKKQBEKEKKREERER-ERERREABLRQOEERTRQOELEALQSKQKA 871  
 Db 1036 RGHLOKRSFQSGISEKQKAEKERALBARAGAGGQQAAGGQVAQGEPRADGG 1095  
 QY 872 ELTRELEKQENQOVELRLKEIEDLOKMEQOELSTFASLOQLQERRODELARLB 931  
 Db 1096 HLASPEVQPSDPSLEHSSPEKAPSPKTLPPQKTVAAESHEKVPSSSEKESRR---1152  
 QY 932 EACRAQOEPLESINPD-EIDECVRNIERSLSGGS-BFSSLEASACEEKPNFQOPYP 988  
 Db 1153 -----QRGLEHVFKQNKHIOCKE--ESALREPSRRVTOEQGVSLBEDK-----K 1195  
 QY 989 EEEVDEGF-----EADDAFK--DSPNPEHGHSDORTSGIRTSDDSSSEEDPYMNDTV 1040

Db 1196 ESREDETLIVTEBAENTSOQOPEQOMAVGKVSSETE--KTLPSGS---PRQOLER 1250  
 QY 1041 PPSPADSTVLLAPSVQDSGLHNSSGESTYCMQONAGDL-SPDGDVDYODDYEDG- 1098  
 Db 1251 PTLALDSRV---SEPPAGSAPETPEDSKPCGSPRVOEKDPSGGSTQIOR--YLDAE 1304  
 QY 1099 -----ALTSGGSVTFPSNSYGSQWSPDYRCVGTYNSSGAVFSSGAGQSPEDSE---1148  
 Db 1305 RLASAVELMRGKVLVAASPSAMLSQSLDS-DRHATGALALPTEERTSTSTSDVSL 1363  
 QY 1149 -----EDFDSRPDTDELSTYRDSYSCVTLPEFHSFLYMKGLMNSWGRKRCVLKDET 1203  
 Db 1364 LPSLAKAOPAEATTGERSAKKPAVOK-----1390  
 QY 1204 LMFRSQOELKQGMHLKKGGSSTL-----SRNKKRMFVLRQSKLMYFEND 1251  
 Db 1391 -----KRGDASLPLDAGLSGSOVDSKSTRK--LFLHKTK-----D 1426  
 QY 1252 SEEKLGIVEVTAKE---IIDNTKENGIDITIMADRTFHLIAE---SPEDASQMFVYL 1304  
 Db 1427 KYYSLEGAELLENVAGHYVLEATTMKGLBAPSGQHRHAGEKRTKEPGKGNRV 1486  
 QY 1305 SOVHASTDOEIOEMDEQANPONAVGTLDVGLIDSVCASDSPRPNSEVITANRVLHN 1364  
 Db 1487 KIGKLTSEKMBESVFRQITNANBELKYLDEFILNKINDLSQKTEPIESLFIATEKFRSN 1546  
 QY 1365 ADT-----PEEMHHWITLL-----QREKGRVBSQGF-----YVGMHLK 1400  
 Db 1547 IKTMVSVNGKIHVGKIDMENYQIVNSNLATERQOKTNLNLFQSLDEFTNGYTKN 1606  
 QY 1401 E---VNNSPKNSLKLKRMFVLTHNS---LDYKSSSEKNALK-----LGTVLNLSLC 1447  
 Db 1607 DEEPVKQS-KAQKKRKQERAVQEHNGHYFASVYQISQSGCGLSYIMWMDKALLCSVC 1665  
 QY 1448 SVPPDEKIFKETGYMNVTYGRKHCYRLYTKLLEATRWSSVIONVDTKAPIDTPQO 1507  
 Db 1666 KMTCHKKCVHKIQSHCSYT-YGRK-----GEPGAPGHFGVCVDSLTSDKASVPVLEK 1718  
 QY 1508 LLODKENCLNSDVVEQYKRN-----PLRYHHPHLS-----1541  
 Db 1719 LLEHVEHMGJL---YBGLYRRSGAANRTRELROALQTPPAVKLEPFILHITGVLYKOWL 1775  
 QY 1542 -----PLPL-PLYGININLNDKGYTLLODEAKIFNSLOQLESMSDPIPIIOGILQY 1595  
 Db 1776 RELPEPLMTFAQYDPL-----RAVELPEKQEQALAI---YAVLEHLPEAN 1818  
 QY 1596 HDL--RPLRDELYCOLIKQTNKVPHPGSVGNLYSQIITC--LSTPLPSRGLKYLKFN 1651  
 Db 1819 HNSLERLIFHLVAKVALLBVDVRM--SPGALAIIFAPCLRCPDNSDPLTSMKDVLEKITTVCV 1877  
 QY 1652 LKRIEOPRGTEMEKALFTYESLKTGCRFEVPSRDEIBALIHQEMTSTVYCHGGSC 1711  
 Db 1878 EMLIKEQ---WKRYKV-KNEEISQLEABASIAR-RLSLRQNAKNSPKTRBAGGAG 1930  
 QY 1712 KITINSHTTAGEVEKILRG 1731  
 Db 1931 RLTTTSRVSPSPSTNNALG 1950

Search completed: July 14, 2003, 18:24:10  
 Job time : 68.7074 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 21.3952 Seconds  
(without alignments)  
3987.653 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCIQGTWLRNQHPSST.....MIVKKYSTRSASSQSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10626	99.0	2058	MY10_HUMAN	Q9h667 homo sapien
2	9961	92.8	2052	MY10_BOVIN	P79114 bos taurus
3	1840.5	17.2	2215	MY7A_MOUSE	P97479 mus musculu
4	1831	17.1	2215	MY7A_HUMAN	Q13402 homo sapien
5	1593	14.8	3511	MY15_MOUSE	Q9qz24 mus musculu
6	1589.5	14.8	3530	MY15_HUMAN	Q9ukn7 homo sapien
7	1504.5	14.0	1846	MY5B_RAT	P70569 ratu
8	1500	14.0	2116	MY5B_DICDI	P08799 dictyostell
9	1492	13.9	2158	MY5B_HUMAN	Q13459 homo sapien
10	1490.5	13.9	2245	MY5B_DICDI	P54697 dictyostell
11	1471	13.7	1980	MY5B_RAT	Q63358 ratu
12	1466	13.7	1855	MY5A_HUMAN	Q9y411 homo sapien
13	1448.5	13.5	1742	MY5C_HUMAN	Q9nqf4 homo sapien
14	1444.5	13.5	1828	MY5A_RAT	Q9qy43 ratu
15	1442	13.4	1574	MY52_YEAST	P19524 saccharomyc
16	1440	13.4	2114	MY51_MOUSE	Q9qy06 mus musculu
17	1434	13.4	1853	MY5A_MOUSE	Q99104 mus musculu
18	1432	13.3	1829	MY5B_CHICK	Q02440 gallu
19	1393.5	13.0	1262	MY0E_HUMAN	Q9un54 homo sapien
20	1373.5	12.8	1509	MY5A_YEAST	P03569 acanthamoeb
21	1362.5	12.7	1471	MY5A_YEAST	P32492 saccharomyc
22	1344	12.5	1978	MY5B_CHICK	P10587 gallu
23	1341.5	12.5	1972	MY5B_RABIT	P35748 cryetolagus
24	1341	12.5	1938	MY5B_CHICK	P13538 gallu
25	1333	12.4	1972	MY5B_HUMAN	P35749 homo sapien
26	1331.5	12.4	1940	MY5B_CHICK	P02565 gallu
27	1323	12.3	2017	MY5B_DROME	Q99323 saccharomyc
28	1322.5	12.3	1928	MY5B_YEAST	Q08638 saccharomyc
29	1320.5	12.3	1972	MY5B_MOUSE	Q08638 mus musculu
30	1320.5	12.3	1976	MY5B_HUMAN	P35580 homo sapien
31	1319.5	12.3	1976	MY5B_BOVIN	Q27991 bos taurus
32	1317.5	12.3	1976	MY5B_RAT	Q9j100 ratu
33	1314	12.2	1941	MY5B_HUMAN	Q9ukx2 homo sapien

34	1313.5	12.2	1938	MY5B_MOUSE	P24733 aequipecten
35	1311	12.2	1937	MY5B_HUMAN	P13535 homo sapien
36	1306.5	12.2	1969	MY5A_CAEL	P12844 caenorhabd
37	1297.5	12.1	1960	MY5B_HUMAN	P35579 homo sapien
38	1294	12.1	1935	MY5B_HUMAN	P12883 homo sapien
39	1288	12.0	1938	MY5D_CAEL	P02567 caenorhabd
40	1287.5	12.0	1935	MY5D_DICDI	P34109 dictyostell
41	1286.5	12.0	1935	MY5B_PIG	P79293 sue scrofa
42	1286	12.0	1966	MY5B_CAEL	P02566 caenorhabd
43	1285	12.0	1938	MY5B_MOUSE	Q02566 mus musculu
44	1279	11.9	1939	MY5B_HUMAN	P12882 homo sapien
45	1278	11.9	1959	MY5B_CHICK	P14105 gallu

## ALIGNMENTS

RESULT 1

MY10\_HUMAN STANDARD; PRT; 2058 AA.

AC Q9HD67; Q9NYM7; Q9P110; Q9UHF6;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Myosin X.

GN MYO10.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.

OX NCBI\_TaxID=9606;

RY [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20442246; PubMed=10984435;

RA Berg J.S., Deffler B.H., Pennisi C.M., Corey D.P., Cheney R.E.

RT "Myosin-X, a novel myosin with pleckstrin homology domains, associates

RL with regions of dynamic actin."

RL J. Cell Sci. 113:3439-3451(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Rogers M.S., Strehler E.E.

RT "Identification of myosin X as a specific binding partner for the

RL tumor suppressive calmodulin-like protein."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA TISSUE=Colorectal adenocarcinoma;

RL Takeda T., Farrell T.J., Anderson J.T., Pourmottabed T.;

RL "Cloning of human myosin X."

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 347-495 FROM N.A.

RA TISSUE=Skeletal muscle;

RL MEDLINE=20079149; PubMed=10610710;

RA Rojas K., Serrano de la Pena L., Gallardo T., Simmons A., Myce K.,

RA McGrath R.J., Considine E., Vasko A.J., Peterson E., Grady D., Cox R.,

RA Andrew L.J., Lovett M., Overhauser J., Williams C.J.;

RT "Physical map and characterization of transcripts in the candidate

RL interval for familial chondroblastoma at chromosome 5p15.1."

RL Genomics 62:177-183(1999).

CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE

CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.

CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS

CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY

CC SIMILARITY). PLAYS A ROLE IN REGIONS OF DYNAMIC ACTIN.

CC -1- SIMILARITY: CONTRAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -1- SIMILARITY: CONTRAINS 2 PH DOMAINS.

CC -1- SIMILARITY: CONTRAINS 2 PH DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; AF247457; AA68025.2; -  
 DR EMBL; AF234532; AA67875.1; -  
 DR EMBL; AF132021; AA636524.1; -  
 DR EMBL; AF132022; AA636525.1; -  
 DR EMBL; AF184153; AA17363.1; -  
 DR HSSP; P08799; LMND.  
 DR Genew; HGNC; 7593; MYO10.  
 DR MIM; 601481; -  
 DR InterPro; IPR000299; Band\_4.1.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR000857; MYTH4.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00663; myosin\_head.  
 DR Pfam; PF00612; IQ; 3.  
 DR Pfam; PF00784; MYTH4; 2.  
 DR Pfam; PF00612; IQ; 3.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR PRODOM; PD000355; myosin\_head; 1.  
 DR SMART; SM00295; B41; 1.  
 DR SMART; SM00015; IQ; 3.  
 DR SMART; SM00242; MYSC; 1.  
 DR SMART; SM00139; MYTH4; 1.  
 DR SMART; SM00233; PH; 2.  
 DR PROSITE; PS00660; BAND\_4.1; FALSE\_NEG.  
 DR PROSITE; PS00661; BAND\_4.1; FALSE\_NEG.  
 DR PROSITE; PS50057; BAND\_4.1; 1.  
 DR PROSITE; PS50096; IQ; 2.  
 DR PROSITE; PS50003; PH\_DOMAIN; 2.  
 DR Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.  
 DR MYOBL; 1; 727  
 FT DOMAIN 1 727  
 FT 742 763  
 FT 764 787  
 FT 788 817  
 FT 1212 1310  
 FT 1392 1497  
 FT 1697 1981  
 FT 799 943  
 FT 98 98  
 FT 148 148  
 FT 256 256  
 FT 324 324  
 FT 1186 1186  
 FT 1663 1663  
 FT 2058 AA; 237389 MW; E4E19BF92D87F3AC CRC64;  
 SQ SEQUENCE

Query Match 99.0%; Score 10626; DB 1; Length 2058;  
 Best local Similarity 99.3%; Pred. No. 0;  
 Matches 2043; Conservative 4; Mismatches 8; Indels 2; Gaps 2;

QY 1 FCLQGRVWLRNGQHPPSTVNSCAEGIVFRTDYQVFTYKOSTITTHQKVTAMHPTNEE 60  
 DB 4 FTEGGRVWLRNGQHPPSTVNSCAEGIVFRTDYQVFTYKOSTITTHQKVTAMHPTNEE 63  
 QY 61 GVDNMASTLHGGSTMYNLFFQYKRNQITWYIGSLASNPPOPLAGLYEPATMYOYR 120  
 DB 64 GVDNMASTLHGGSTMYNLFFQYKRNQITWYIGSLASNPPOPLAGLYEPATMYOYR 123  
 QY 121 RHILGELPHIFALANECYRCLMRHNOCTILKSGSAGKTESKTLILKFLSVISQOSLE 180  
 DB 124 RHILGELPHIFALANECYRCLMRHNOCTILKSGSAGKTESKTLILKFLSVISQOSLE 183  
 QY 181 LSLIKETSCVERAIISSPIMEAFGNAKTYNNNSRFGKTYOLNTCQKNGIQQGRIYVC 240  
 DB 184 LSLIKETSCVERAIISSPIMEAFGNAKTYNNNSRFGKTYOLNTCQKNGIQQGRIYVC 243  
 QY 241 ILSNORVVRNGENNYHIFALLAGLEHEEREETLSPENYHHLNOSGVEDKTID 300  
 DB 244 LL-EKNRVVQNGENNYHIFALLAGLEHEEREETLSPENYHHLNOSGVEDKTID 302

QY 301 QESFREVITAMVMOFSKEVEVSRLLAGILHLNIBETTAGAQVSEKTLGSRABLL 360  
 DB 303 QESFREVITAMVMOFSKEVEVSRLLAGILHLNIBETTAGAQVSEKTLGSRABLL 362  
 QY 361 GLDPTQLTDALTQSRMFLRGEELITPLNAVQAQVDSRDLAALVACCEBWTIKINSRIK 420  
 DB 363 GLDPTQLTDALTQSRMFLRGEELITPLNAVQAQVDSRDLAALVACCEBWTIKINSRIK 422  
 QY 421 GNEDFKSIGILDIQFENPEVNHFPQFNINYNEXLQEFYNGHIFSLQLEYSRGLVWE 480  
 DB 423 GNEDFKSIGILDIQFENPEVNHFPQFNINYNEXLQEFYNGHIFSLQLEYSRGLVWE 482  
 QY 481 DIDWIDNGECDLIEKKLGLLALINEESHFPQATSTLLEKLSQANHHFYVKEPRVAVN 540  
 DB 483 DIDWIDNGECDLIEKKLGLLALINEESHFPQATSTLLEKLSQANHHFYVKEPRVAVN 542  
 QY 541 NFGVHXAGEVOYDVRCILKRNKRDTPRDDLNLBESRPFYDLPEHYSSNNODTLKC 600  
 DB 543 NFGVHXAGEVOYDVRCILKRNKRDTPRDDLNLBESRPFYDLPEHYSSNNODTLKC 602  
 QY 601 GSKHRRPTVSSQFKVDSLHSMATLSSNPFVRCIKPMQKMPQFDQAVVNLQRYSG 660  
 DB 603 GSKHRRPTVSSQFKVDSLHSMATLSSNPFVRCIKPMQKMPQFDQAVVNLQRYSG 661  
 QY 661 MLETVRIRKAGYAVRPFQDFYKRYKVMRNALBEDVNGKCTSLLOLYDASNSEMOIGK 720  
 DB 663 MLETVRIRKAGYAVRPFQDFYKRYKVMRNALBEDVNGKCTSLLOLYDASNSEMOIGK 721  
 QY 721 TVYFLRESLEQLEKREBEVSHAMVIRAHVGLGLAKQRYKVLVCYVITQKNTAFLL 780  
 DB 722 TVYFLRESLEQLEKREBEVSHAMVIRAHVGLGLAKQRYKVLVCYVITQKNTAFLL 781  
 QY 781 RRRFLHLKKAALVFOKOLRGQIARVYROLAEKKEOEKKOEKKEEKKREERERER 840  
 DB 782 RRRFLHLKKAALVFOKOLRGQIARVYROLAEKKEOEKKOEKKEEKKREERERER 841  
 QY 841 ERREBELRAQOEESTRKOQELBALOKSKEALATRELEKOKENKQVEILRLKEIEDLO 900  
 DB 842 ERREBELRAQOEESTRKOQELBALOKSKEALATRELEKOKENKQVEILRLKEIEDLO 901  
 QY 901 RKKEQOEISLTASIQLOERRDQELRLBESACPAQOFLSLNPFDEICVRNIERSL 960  
 DB 902 RKKEQOEISLTASIQLOERRDQELRLBESACPAQOFLSLNPFDEICVRNIERSL 961  
 QY 961 SGSGSFSSSLASACEBKPNFNSQPYEBEVEDEBADDPAFKDSPNPSSEHGHSDQRTS 1020  
 DB 962 SGSGSFSSSLASACEBKPNFNSQPYEBEVEDEBADDPAFKDSPNPSSEHGHSDQRTS 1021  
 QY 1021 GRTSDSDSEEDPYNNDTVPSPADSTVLLAPSVODSGSLHNSSGESTYCMQONAGD 1080  
 DB 1022 GRTSDSDSEEDPYNNDTVPSPADSTVLLAPSVODSGSLHNSSGESTYCMQONAGD 1081  
 QY 1081 LPSPPGDYDODDYEDGALITSGSVTPNSNSGQSPEDYRCSVGTYNSSGAYRPSBEA 1140  
 DB 1082 LPSPPGDYDODDYEDGALITSGSVTPNSNSGQSPEDYRCSVGTYNSSGAYRPSBEA 1141  
 QY 1141 OSSFEDSEEDPSRDTDELISYRDSVYSCVTLFPHSFLYMKGGLNMSGRKRCVLYKD 1200  
 DB 1142 OSSFEDSEEDPSRDTDELISYRDSVYSCVTLFPHSFLYMKGGLNMSGRKRCVLYKD 1201  
 QY 1201 EFTLWFRSROEALOKGWLHKKGGSSTLSRRMKRWFLRQSKLMEFENDESEKLGKTV 1260  
 DB 1202 EFTLWFRSROEALOKGWLHKKGGSSTLSRRMKRWFLRQSKLMEFENDESEKLGKTV 1261  
 QY 1261 EYRTAKEIITDNTKENGIDITMAADRTPHILASPEDASQWVSQVASTQOEIOENHD 1320  
 DB 1262 EYRTAKEIITDNTKENGIDITMAADRTPHILASPEDASQWVSQVASTQOEIOENHD 1321  
 QY 1321 EOPANQNAVGLTDVGLIDSVSCASDSPDRPNSFVITANRVLHCNADTPEBMMHWTTLQOR 1380  
 DB 1322 EOPANQNAVGLTDVGLIDSVSCASDSPDRPNSFVITANRVLHCNADTPEBMMHWTTLQOR 1381



QY	301	QSSPFEVITAMVMOSESKEVEVSSLAGIIGHIERITTCGAGOVSEKTLGSAEILL	360
Db	303	QSSPFEVITAMVMOSESKEVEVSSLAGIIGHIERITTCGAGOVSEKTLGSAEILL	362
QY	361	GLDPQOLTDALQORSNFWLGESEILLPLNVOQAVDSHDSJAMALYACCPEFVILIKKINSRIK	420
Db	363	GLDPQOLTDALQORSNFWLGESEILLPLNVOQAVDSHDSJAMALYACCPEFVILIKKINSRIK	422
QY	421	GNEDFKSICILDI FGFEENFEVNHFEOPFNINYANEKLOSTYFNKHI FSELEOLEYSREGULWE	480
Db	423	GNEDFKSICILDI FGFEENFEVNHFEOPFNINYANEKLOSTYFNKHI FSELEOLEYSREGULWE	482
QY	481	DI DMI DNGECCLDIEKKGLLALINBESHFPQATOSTTLEKSHOHANNHPYKPPVANA	540
Db	483	DI DMI DNGECCLDIEKKGLLALINBESHFPQATOSTTLEKSHOHANNHPYKPPVANA	542
QY	541	NFGVGHYAGEVOYDVARGLIEKRNDFPRDILLNLBRESRDFIYDLFEHVSRRNQDTLKC	600
Db	543	NFGVGHYAGEVOYDVARGLIEKRNDFPRDILLNLBRESRDFIYDLFEHVSRRNQDTLKC	602
QY	601	GSKHRRPYTSSQFKVDSLSHLMATLSSSNPFYRCIKPMNQKMPDOPDOAVLNQRLRYSG	660
Db	603	GSKHRRPYTSSQFK -DSLHLMATLSSSNPFYRCIKPMNQKMPDOPDOAVLNQRLRYSG	661
QY	661	MLETRIRIRAGAVNRPDPFYKRYVULMRNIALPBDVAGKCSLSLOLDASNSEWOLK	720
Db	662	MLETRIRIRAGAVNRPDPFYKRYVULMRNIALPBDIGKCTALLOLDASNSEWOLK	721
QY	721	TKVFLPRESLEQKLKREBEVSHAAVNIAPHVGLFARQOYRKVLYCWIYLOKNYRAPFL	780
Db	722	TKVFLPRESLEQKLKREBEVTRAAMVNIAPHVGLYARQOYKVLDCWIIYLOKNYRAPFL	781
QY	781	RRRFLHLKKAALVFPQQLRGGOTARRPYRQLLAKRQOEKKKQOESEKKRBESEERER	840
Db	782	RRRFLHLKKAALVFPQQLRGGOTARRPYRQLLAKRQOEKKRBESEKKRBESEERER	841
QY	841	ERRAEALRQOESEETKQOELEAL -OKSOEALTELEKOKENKOVEETILREKETIEDL	899
Db	842	ERRAEALRQOESEETKQOELEALQOESQOBAEELSLEKOKENKOVEETILREKETIEDL	901
QY	900	QRMKSEQELSTLEASLOKQOERRDOELRLLEEBACRAQEPLESLNFDIEDECVRNIERS	959
Db	902	QRMKSEQELSTLEASLOKQOERRDEBRLLEBACRAQEPLESLNFDIEDECVRNIERS	961
QY	960	LSGSGEFSSELASACEKFNPNFSQPY -EEVYDGEFEADDAFODSPNPSHGHSDDR	101
Db	962	LSVGGGCTGE -OGAQAERPSFNFSQPYDEEBEVBGEFEADDAFODSPNPSHGHSDDR	101
QY	1019	TSGITRSDSSSEEDPYMNDTVVPTSPSASSTVLLASVODSGSLHNHSSGESYTCMPONA	1077
Db	1020	TSGITRSDSSSEEDPYMNDTVVPTSPSASSTVLLASVODSGSLHNHSSGESYTCMPONA	1077
QY	1079	GOLPSPDGDYDODDYEDGAIITSGSVTFSNSYGSQWSPDYKCSVGTYNSSGAYRFSSE	113
Db	1074	GALPAPBEGDYDODDYEDGAIITSGSVTFSNSCSQWSPDYKCSVGTYNSSGAYRFSSE	113
QY	1139	GAQSSPEDEBDFDFTDDDELSTYRDSVYSCVTLLPYTHSPLYMNGGLMNMKRWCTL	119
Db	1134	GAQSSPEDEBDFDFTDDDELSTYRDSVYSCVTLLPYTHSPLYMNGGLMNMKRWCTL	119
QY	1139	KOETFLMFSKOBALKQGLJHKKGSSSTLSRNMKGMFVLRQSLKMTFENDSEKTLG	125
Db	1134	KOETFLMFSKOBALKQGLJHKKGSSSTLSRNMKGMFVLRQALKMTFENDSEKTLG	125
QY	1239	TVEVTRAKEDIDNTTKENGIDIIIMADRTFHLIABSPEDASQWFSVLSOVHASTDOEIOEM	131
Db	1254	TVEVTRAKEDIDNTTKENGIDIIIMADRTFHLIABSPEDASQWFSVLSOVHASTDOEIOEM	131
QY	1319	HDEQANPONAUGTLDVGLIDSYCASGSPRRPNFSVITIANRVLHCNADTPREEMHMITLL	137
Db	1314	HDEQANPONAUGTLDVGLIDSYCASGSPRRPNFSVITIANRVLHCNADTPREEMHMITLL	137

QY	1379	ORSGDPRVBEQSEITVGMHLKHEKYNPKPKSSLT,KRWFWLTNNSLDYKSSSKNAL,KL	1438
Db	1374	QRSGKDRVBEQSEITVGMHLKHEKYNPKPKSSLT,KRWFWLTNNSLDYKSSSKNAL,KL	1433
QY	1439	GTLVNLSLCSVVPDEKIFKETGYMANTVYGRKHCYBL,YTKLNEATRMSSVIONTDTK	1498
Db	1434	GTLVNLSLCSVVPDEKIFKETGYMANTVYGRKHCYBL,YTKLNEATRMSVAIONVDTK	1493
QY	1499	APIPTPTQOLI,ODIKENCLNSADV,EOIYKKNPILRTYHHPLHSLPL,PPYGDINLNL,KD	1558
Db	1494	APIPTPTQOLI,ODIKENCLNSADV,EOIYKKNPILRHTHHP,HLHSLPL,PPYGDINLNL,KD	1553
QY	1559	KGYTTLODEAKITNSLQOLESMDP,PIITIGILQTHD,LRLP,LEDEL,YCGLIKOTNV,PH	1618
Db	1554	KGYTTLODEAKITNSLQOLESMDP,PIITIGILQTHD,LRLP,LEDEL,YCGLIKOTNV,PH	1613
QY	1619	PGSVGNLSWMOILTCLSTCLP,PSRGI,KYL,KFHLKIR,IREOP,PGTEMEKVAL,FTYESL,KT	1678
Db	1614	PGSVGNLSWMOILTCLSTCLP,PSRGI,KYL,KFHLKIR,IREOP,PGTEMEKVAL,FTYESL,KT	1673
QY	1679	KCREFWPSRDEIEAL,IRHOEMTSTVYCHGGSGCKITINSHTAGEVVEKLI,IRGLAMEDSR	1738
Db	1674	KCREFWPSRDEIEAL,IRHOEMTSTVYCHGGSGCKITINSHTAGEVVEKLI,IRGLAMEDSR	1733
QY	1739	NMFALF,FEVNGVDAIKASIR,TVAAVLAK,FEKLTATS,EGD,PMKF,FKYLC,CF,LD,TVNPK	1798
Db	1734	NMFALF,FEVNGVDAIKASIR,TVAAVLAK,FEKLTATS,EGD,PMKF,FKYLC,CF,LD,TVNPK	1793
QY	1799	DSVEFAFMFOEQAHEAV,IGHHP,PAEENI,OV,LAAL,RLQYLOGDYTL,HLAIP,LEEV,SLOR	1858
Db	1794	DSVEFAFMFOEQAHEAV,IGHHP,PAEENI,OV,LAAL,RLQYLOGDYTL,HLAIP,LEEV,SLOR	1853
QY	1859	LKARISOSTKTF,TCERLEKRRIS,FLBGLT,RRSFR,TSVVRQV,KEEOM,LD,MM,IKEEV,SS	1918
Db	1854	LKARISOSTKTF,TCERLEKRRIS,FLBGLT,RRSFR,TSVVRQV,KEEOM,LD,MM,IKEEV,SS	1913
QY	1919	ARASITIDMKR,FCOQMNQO,AMAKY,MAI,IKEMPG,SGTL,FDVECKEGG,POELM,IGV,ADA	1978
Db	1914	ARASITIDMKR,FCOQMNQO,AMAKY,MAI,IKEMPG,SGTL,FDVECKEGG,POELM,IGV,ADA	1973
QY	1979	VSVYKRGEGR,LEFYQYEH,ILSP,APANTYK,IV,DERELL,FT,SEV,VDVAK,LMKAYISM	2038
Db	1974	VSVYKRGEGR,LEFYQYEH,ILSP,APANTYK,IV,DERELL,FT,SEV,VDVAK,LMKAYISM	2033
QY	2039	IVKKRYSTTRSASSQSSSR 2057	
Db	2034	IVKKRYSTRSRVSSQSSSR 2052	
RESULT 3			
MY7A_MOUSE			
ID	MY7A_MOUSE	STANDARD;	PRT; 2215 AA.
AC	P97479;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DB	Myosin VIIa.		
GN	MYO7A OR MYO7.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95174881; PubMed=7870172;		
RA	Gibson F., Walsh J., Mburu P., Varrelle A., Brown K.A., Antonio M.,		
RA	Beisel K.W., Steel K.P., Brown S.D.M.;		
RT	"A type VII myosin encoded by the mouse deafness gene shaker-1.";		
RL	Nature 374:62-64(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96343888; PubMed=9680294;		
RA	Mburu P., Liu X.-Z., Walsh J., Saw D. Jr., Cope M.J., Gibson F.,		



Db 821 QACRAVLVKAERHRLMAVITVOAYARGLIARLHRLRLEVEYORLRLAEERMLAEK 880  
 Qy 830 KRE-----EEREREREREAELRAQOEETRKOELEALOKOEALRELEKOK-- 881  
 Db 881 RKMMAKKAKEERERGERLQALAREDAE-----RELKEERARKKELLEMEKARHE 935  
 Qy 882 ---ENKQVEIL-----RLEKEIEDLORMKEQOELSLEASLOKLQERDO 924  
 Db 936 PINHSMVDKMFPGFLTSGSLPGQEQAPGCFEDLERGR----- 974  
 Qy 925 ELARLEECRAQOEFLESINPEIDECVNIERSLSGSEFSESLAESACEKPNPNS 984  
 Db 975 ---REWEEDVDALPLP---DEDEDSEYFAFAATYFQCTTHS--YTRP--LK 1023  
 Qy 985 QP---YEEBEYDE-----GFEADDAFDPSNPSEHSGDQRTSGIRTSDDS 1028  
 Db 1024 QPLHYHDBEDQALALAVITLIRFG-----DLPERKH-----TMSDG 1064  
 Qy 1029 SEEDPYM-----NDIVPTSPSADST--VLLAPSVODSGSLHN 1064  
 Db 1065 SEKIPLYMTKIYETLGGKTYKRELQALQGEGETQLPEGQKTSYRHKLVHLTLKKKSLTE 1124  
 Qy 1065 S-----SSGESTYCMQNGDLPSP-----GDYDYOQDDVEDDAIS 1102  
 Db 1125 EYTKRLNDGESTYQGSNMLEDRPTSNLEKHLIINGILRALRDLITCYQISKQ--LTH 1181  
 Qy 1103 GSSVTFNSYSGSQSPDYRCSTVGTYNSS-----GAYRFSSEGAQGSFED 1146  
 Db 1182 NPS---KSSYARGMILVSLC--VGCFAPESEKFKYKLNFIHGPGYAPCEBELRRTYFN 1237  
 Qy 1147 SEEDFPRFPTDELSTYRDSYVSCTLYFHSFLYMKGGMLN-----SWKRWG--- 1196  
 Db 1238 G-----TRTOPPSMLEQATSKSKPIMLP---VTWMDGTTKTLTDSATTARELCNALA 1288  
 Qy 1197 ---VLDD---ETFLMFRSKQELKOGMLH-----KKGSSSTLSRMKKRM 1237  
 Db 1289 DKISLDRGFSLYIALFDYVSSLSGSDHMDAISQCEQYAEQQA---QERNAPMWL 1344  
 Qy 1238 FVLROKLMFYENSEEKLKGTV-----EVRTAKE--IIDNTTKENGIDI-- 1280  
 Db 1345 F-PRKVFPRMHPNSPDENVATNLIYQGVYGVKFGYRCKEDDLAELASQYFVYGSB 1403  
 Qy 1281 IMADRTHLIAE-----SP-EDASQWFSVLQVH-----ASTDEIQEMDEQANPON 1327  
 Db 1404 MLERLISLVPTIPIPREITPLKMLEKMAQALAAKKGIVARQRTDSQVXEDVNVYAR 1463  
 Qy 1328 AVGTLDVGLIDSCASDSPRNSPILITANRVLHNCADPREMHMITLQSKGDYV 1387  
 Db 1464 FKMPFLFSRFEYAYKFSGPPLPKSDVIVAVNMTGVYFVDEQBV---LLELSPEEIMA 1518  
 Qy 1388 ---EGQEFIV-----RGWLHKEVKNSPKMS--SLKXKKRMFVLTNHSID 1426  
 Db 1519 VSSSRRCRVLISGSDLCATQOSRAGLTPAGPSCPCSCGTGMAMPFLATIKD 1578  
 Qy 1427 YKSSSEKALKGLTVLNSLCSVPPDEKTFKGTGW---NVTYGRKHCRYLYTK-- 1479  
 Db 1579 EYFTFSNMEDRIDLVVTFP---EGLRKRSKTVVALQDNPAGESGFLSPAKGD 1631  
 Qy 1480 ---LLEAT-----RMSVIONVTDKAPIDT-----PQOQL----- 1509  
 Db 1632 LIILHDHTEQVWNSGANGINERTKORGDFPTDCYVMEPTVTLPREILAVTMTDOR 1691  
 Qy 1510 ODI-----KENCNSDVVEOIIYKRNPIRTYTHP 1538  
 Db 1692 QDVVRLQJRTAPEVRAKPYTLEFSYDFRPPPHITLSRVVNSAKRGDRLMSHTREP 1751  
 Qy 1539 LHSPLPLPYGDINMLLKDGTYTLODEAKIFNSL-----QOLESMDPIPIIQ 1589  
 Db 1752 LKQALLK-----KILGSEESQEAQMAFVAVLKMGDYPSKRMASVNE---LND 1797  
 Qy 1590 GILQTHDLRLADELYCOLIQ--TNKVPHPSGVNLSWOJLTGLSTCFPLSRGILKX 1647  
 Db 1798 QIFEMALKAEPLKDEAVVOILKQLTDNHIRYSEERG---WELLMLCTGLFPSPNILLPH 1853

Qy 1648 LKFHLKRIEOPFGTEMERYALFTYESLKK---TKCRFVPSRDEIALIHR-QEMTSIV 1703  
 Db 1854 VORFLQGS-RKICP-----LAIDCLOQLKALRNSRKRPPLVVEVAIQHTTQIPHKV 1906  
 Qy 1704 YCHGSSCKITINSHTTAGEVVEKILRGLAMEDSRNMAL-----FEY 1746  
 Db 1907 YPPDDTDEAFEVESSTKADPCONTASRLKSSSG-FSLPVKIADKVISVENDFFPDF 1965  
 Qy 1747 NGHVDKAIERSRVVADVLAKEFLAATSEVGDLPKPFYF--KLUCFLDTPNP-RDSV-E 1802  
 Db 1966 VHLHDIMKAPRIDGI-----VBSLITVQVFPMKLM---TTVPGADPMND 2010  
 Qy 1803 FAFMEQAEAVIHGHHPAPEENLOVAL--RLQYLOQDYTLHAAPLEEVYSLORLK 1860  
 Db 2011 SIFHYQELPKYLRGVHKCTREEVLOGLALIVKXF-BEDKSYFPSIPKL----- 2059  
 Qy 1861 ARISOSTKTFTECELEKRRISFLEGTLRSRSTSSVVRQKVEEQMLDMYKEVSSAR 1920  
 Db 2060 -----LR-----ELVPQDLIRQVSPDW-----K 2078  
 Qy 1921 ASIIDKMRKFOGANOQOMAKYMALIKEMPGYSTLFDV-EEKEGFPOLMIGVSADAV 1979  
 Db 2079 KSIIVAFNKHAKSKSEAKLAFKLIKMPFTGSAFVEVKQTTBERPFILIAINKGV 2138  
 Qy 1980 SVYKSGEGRPLEVFQYEHILSPGAPLANTYKIV-----VDERELLFETSEVVDAKLMKA 2034  
 Db 2139 SLIDRPTKOILTTHPTKISNWS--GNTYFHTITGNLVRSGLKLCETSLGKMDLLTS 2196  
 Qy 2035 YISMVYKRYSTTRASASSGSSR 2057  
 Db 2197 YISQWL-----TAMSKORNSR 2212

RESULT 4  
 MY7A HUMAN  
 ID MY7A HUMAN STANDARD; PRT; 2215 AA.  
 AC Q13402; Q92821; Q92822; Q13321; P78427; Q14785;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin VIIa.  
 GN MYO7A OR USH1B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1, 5, 6 AND 7), AND DEVELOPMENTAL  
 RP EXPRESSION.  
 RC TISSUE=Retina;  
 RX MEDLINE=96194949; PubMed=8622919;  
 RA Weil D., Levy G., Sahly I., Levi-Acobas F., Blanchard S.,  
 RA El-Amraoui A., Crozet F., Philippe H., Abitbol M., Petit C.,  
 RT "Human myosin VIIa responsible for the usher 1B syndrome: a predicted  
 RT membrane-associated motor protein expressed in developing sensory  
 RT epithelia".  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3232-3237(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 4).  
 RC TISSUE=Testis;  
 RX MEDLINE=97038686; PubMed=8884267;  
 RA Chen Z.-Y., Hasson T., Kelley P.M., Schwender B.J., Schwartz M.F.,  
 RA Ramakrishnan M., Kimberling W.J., Mooseker M.S., Corey D.P.;  
 RT "Molecular cloning and domain structure of human myosin-VIa, the gene  
 RT product defective in usher syndrome 1B".  
 RL Genomics 36:440-448(1996).  
 RN [3]  
 RP SEQUENCE OF 166-196 FROM N.A.  
 RC TISSUE=Epithelium, Leukocyte, and Liver;  
 RX MEDLINE=94294418; PubMed=8022818;  
 RA Bement M.M., Hasson T., Wirth J.A., Cheney R.E., Mooseker M.S.;  
 RT "Identification and overlapping expression of multiple unconventional

RT myosin genes in vertebrate cell types.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 91:6549-6553(1994).  
 RN [4]  
 RP SEQUENCE OF 1-1075 FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=96003870; PubMed=7568224;  
 RA Hassan T., Heintzelman W.B., Santos-Sacchi J., Corey D.P.,  
 RT Mooseker M.S.;  
 RT "Expression in cochlea and retina of myosin VIIa, the gene product  
 RT defective in Usher syndrome type 1B.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9815-9819(1995).  
 RN [5]  
 RP SEQUENCE OF 96-564 FROM N.A. (ISOFORM 1), AND VARIANTS USH1B.  
 RC TISSUE=Retina;  
 RX MEDLINE=95174880; PubMed=7870171;  
 RA Weil D., Blanchard S., Kaplan J., Guilford P., Gibson P., Walsh J.,  
 RA Mburu P., Varela A., LeVilliers J., Weston M.D., Kelley P.M.,  
 RA Kimberling W.J., Wagenaar M., Levi-Acobas F., Larget-Biet D.,  
 RA Munnich A., Steel K.P., Brown S.D.M., Petit C.;  
 RT "Defective myosin VIIa gene responsible for Usher syndrome type 1B.";  
 RL Nature 374:60-61(1995).  
 RN [6]  
 RP SEQUENCE OF 79-578 FROM N.A.  
 RX MEDLINE=97224487; PubMed=9070921;  
 RA Kelley P.M., Weston M.D., Chen Z.-Y., Orten D.J., Hassan T.,  
 RA Overbeck L.D., Plint J., Talmadge C.B., Ing P., Mooseker M.S.,  
 RA Corey D.P., Sumegi J., Kimberling W.J.;  
 RT "The genomic structure of the gene defective in Usher syndrome type 1B  
 RT (MYO7A).";  
 RL Genomics 40:73-79(1997).  
 RN [7]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=96440426; PubMed=8842737;  
 RA El-Amraoui A., Sally I., Picard S., Sahel J., Altbol M., Petit C.;  
 RT "Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy  
 RT explained by the presence/absence of myosin VIIa in the photoreceptor  
 RT cells.";  
 RL Hum. Mol. Genet. 5:1171-1178(1996).  
 RN [8]  
 RP VARIANTS USH1B H-212; C-212; H-302; Q-468 INS; L-503 AND Q-450.  
 RX MEDLINE=97055890; PubMed=8900236;  
 RA Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,  
 RA Hassan T., Chen Z.-Y., Corey D.P., Mooseker M.S., Sumegi J.,  
 RA Cremer C., Moeller C., Jacobson S.G., Gorin M.E., Kimberling W.J.;  
 RT "Myosin VIIa mutation screening in 189 Usher syndrome type 1  
 RT patients.";  
 RL Am. J. Hum. Genet. 59:1074-1083(1996).  
 RN [9]  
 RP VARIANTS USH1B ARG-214; ASP-397 AND THR-826, AND POLYMORPHISMS.  
 RX MEDLINE=98018437; PubMed=9382091;  
 RA Adato A., Weil D., Kalinski H., Pel-Or Y., Ayadi H., Petit C.,  
 RA Korostoffsky M., Bonne-Tamir B.;  
 RT "Mutation profile of all 49 exons of the human myosin VIIa gene, and  
 RT haplotype analysis, in Usher 1B families from diverse origins.";  
 RL Am. J. Hum. Genet. 61:813-821(1997).  
 RN [10]  
 RP VARIANTS USH1B ARG-25; SER-955 AND GLU-2137, AND POLYMORPHISMS.  
 RX MEDLINE=97156218; PubMed=9002678;  
 RA Levy G., Levi-Acobas F., Blanchard S., Gerber S., Larget-Biet D.,  
 RA Chapal V., Liu X.-Z., Newton V., Steel K.P., Brown S.D.M., Munnich A.,  
 RA Kaplan J., Petit C., Weil D.;  
 RT "Myosin VIIa gene: heterogeneity of the mutations responsible for  
 RT Usher syndrome type 1B.";  
 RL Hum. Mol. Genet. 6:111-116(1997).  
 RN [11]  
 RP VARIANT DFN2 PRO-244.  
 RX MEDLINE=97315827; PubMed=9171833;  
 RA Liu X.-Z., Walsh J., Mburu P., Kendrick-Jones J., Cope M.J.,  
 RA Steel K.P., Brown S.D.M.;  
 RT "Mutations in the myosin VIIa gene cause non-syndromic recessive  
 RT deafness.";  
 RL Nat. Genet. 16:188-190(1997).  
 RN [12]

RP VARIANT DFN2 ILB-599.  
 RX MEDLINE=97315828; PubMed=9171833;  
 RA Weil D., Kuesel P., Blanchard S., Levy G., Levi-Acobas F., Drita M.,  
 RA Ayadi H., Petit C.;  
 RT "The autosomal recessive isolated deafness, DFN2, and the Usher 1B  
 RT syndrome are allelic defects of the myosin-VIIa gene.";  
 RL Nat. Genet. 16:191-193(1997).  
 RN [13]  
 RP VARIANT DFN11 886-ALA-LYS-888 DEL.  
 RX MEDLINE=98016404; PubMed=9354784;  
 RA Liu X.-Z., Walsh J., Tamagawa Y., Kitamura K., Nishizawa M.,  
 RA Steel K.P., Brown S.D.M.;  
 RT "Autosomal dominant non-syndromic deafness caused by a mutation in the  
 RT myosin VIIa gene.";  
 RL Nat. Genet. 17:268-269(1997).  
 RN [14]  
 RP VARIANTS USH1B PRO-651 AND GLN-1602.  
 RX MEDLINE=98386090; PubMed=9718356;  
 RA Liu X.-Z., Hope C., Walsh J., Newton V., Ke X.M., Liang C.Y., Xu L.R.,  
 RA Zhou J.M., Trump D., Steel K.P., Bunday S., Brown S.D.M.;  
 RT "Mutations in the myosin VIIa gene cause a wide phenotypic spectrum,  
 RT including atypical Usher syndrome.";  
 RL Am. J. Hum. Genet. 63:909-912(1998).  
 RN [15]  
 RP VARIANT USH1B PRO-1087.  
 RX MEDLINE=99294608; PubMed=10364543;  
 RA Adato A., Kalinski H., Weil D., Chab H., Korostoffsky M.,  
 RA Bonne-Tamir B.;  
 RT "Possible interaction between USH1B and USH3 gene products as implied  
 RT by apparent digenic deafness inheritance.";  
 RL Am. J. Hum. Genet. 65:261-265(1999).  
 RN [16]  
 RP VARIANTS USH1B, AND POLYMORPHISMS.  
 RX MEDLINE=99140257; PubMed=10094549;  
 RA Janacke A.R., Meins M., Sadeghi M., Grundmann K., Apfelstedt-Sylla E.,  
 RA Zrenner E., Rosenberg T., Gal A.;  
 RT "Twelve novel myosin VIIa mutations in 34 patients with Usher syndrome  
 RT type I: confirmation of genetic heterogeneity.";  
 RL Hum. Mutat. 13:133-140(1999).  
 RN [17]  
 RP VARIANTS USH1B LYS-1170 AND CYS-1719.  
 RX MEDLINE=99374392; PubMed=10447383;  
 RA Cuevas J.M., Espino C., Millan J.M., Sanchez F., Trujillo M.J.,  
 RA Ayuso C., Benito M., Najera C.;  
 RT "Identification of three novel mutations in the MYO7A gene.";  
 RL Hum. Mutat. 14:181-181(1999).  
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN  
 CC RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON-  
 CC SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS  
 CC DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE RIGIDITY OF STEREOCILIA  
 CC DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE. INVOLVED IN HAIR-CELL  
 CC VESICLE TRAFFICKING OF AMINOGLYCOSIDES, WHICH ARE KNOWN TO INDUCE  
 CC OTOTOXICITY (BY SIMILARITY).  
 CC -1- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE  
 CC FORMATION OF A COILED-COIL ROD.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE). IN THE PHOTORECEPTOR  
 CC CELLS, MAINLY LOCALIZED IN THE INNER AND BASE OF OUTER SEGMENTS AS  
 CC WELL AS IN THE SYNAPTIC ENDING REGION.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; 1 (SHOWN HERE), 2, 3,  
 CC 4, 5, 6 AND 7; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PIGMENT EPITHELIUM AND THE  
 CC PHOTORECEPTOR CELLS OF THE RETINA. ALSO FOUND IN KIDNEY, LIVER,  
 CC TESTIS, COCHLEA, LYMPHOCYTES. NOT EXPRESSED IN BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: DETECTED IN OPTIC CUP IN 5.5 WEEKS-OLD  
 CC EMBRYOS. EXPRESSED IN RETINAL PIGMENT EPITHELIUM, COCHLEAR AND  
 CC VESTIBULAR NEUROEPITHELIUM, AND OLFACTORY EPITHELIUM AT 8 WEEKS. AT  
 CC 19 WEEKS, PRESENT IN BOTH PIGMENT EPITHELIUM AND PHOTORECEPTOR  
 CC CELLS. AT 24-28 WEEKS, EXPRESSION IN PIGMENT EPITHELIUM AND  
 CC PHOTORECEPTOR CELLS INCREASES. PRESENT IN PIGMENT EPITHELIUM AND  
 CC PHOTORECEPTOR CELLS IN ADULT.



CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF USHER SYNDROME TYPE 1B  
CC (USH1B), ALSO KNOWN AS USHER SYNDROME NON-ACADIAN VARIETY. USHER  
CC SYNDROME TYPE 1 IS AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY  
CC PROFOUND CONGENITAL SENSORINEURAL DEAFNESS, CONSTANT VESTIBULAR  
CC DYSFUNCTION AND PREPUBERTAL ONSET OF PROGRESSIVE RETINITIS  
CC PIGMENTOSA LEADING TO BLINDNESS. USHER SYNDROME IS THE MOST COMMON  
CC CAUSE OF COMBINED DEAFNESS AND BLINDNESS IN DEVELOPED COUNTRIES.  
CC USH1A APPEARS AS A PRIMARY CYTOSKELETAL PROTEIN DEFECT.  
CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE  
CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB2 OR NSRD2) WITH  
CC VARIABLE VESTIBULAR DYSFUNCTION AND VARIABLE AGE OF ONSET.  
CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL DOMINANT  
CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNALL) WITH ONSET  
CC AFTER COMPLETE SPEECH ACQUISITION AND SUBSEQUENT GRADUAL  
CC PROGRESSION.  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 5 IQ DOMAINS.

Query Match 17.1%; Score 1831; DB 1; Length 2215;  
Best Local Similarity 26.1%; Pred. No. 2e-76;  
Matches 632; Conservative 379; Mismatches 822; Indels 592; Gaps 87;

QY 4 QGTRVWL-RENGQHPSTVSCAGIYVFTDGVFTYKOSTIT- - - - -QKTAH 55  
DB 6 QGHVWMDLRLGDFVPI- - - - -GAVVKJCSGQVQVDEDENEHISQONATHIKPMH 60  
QY 56 PTNEGVDMASLTSLHGSIMYNLFOYRKNQIWTYISILASVNPYOPIALYEPATM 115  
DB 61 PTVHGVEDMIRGDLNEAGILNLIRYDHLIYTSGLIIVAVPYO-LLSIYEPHI 119  
QY 116 EYSRRLGELPHIFAIANECEYRCLMKRHDNOCILIKESGAGKTESYLIIKPLSVIS 175  
DB 120 RQYTNKIGMPHIFAIADNCYFNMKRNSRDOCCIISGSGAGKTESYLIIQPLAIS 179  
QY 176 QOSLESLAKETCSVERALIESPIMEAFGNATVNNNSRFGKFPQULNICOKNIOG 235  
DB 180 GQH- - - - -SWIEQVLEATPILEAFGNATKIRNDSSRFGKIDIHFNRGALIEGA 230  
QY 236 RIYDCILISQNRVVRQNGERNYHIFAYALAGLEHEHEEFYLTSTENYHILNOSGVED 295  
DB 231 KIQOYL- EKSRCVROLDERNHVFCMLGMSBQKKLGIGASDVNYIYLMGNCITC 289  
QY 296 KITSDOESPREVITANDMOPSKBEVREVSRLIAGILHGNIE- - - - -FTAGCAOVSEK 350  
DB 290 EGVDSOEVANINSAMKVLMTFTDENWELSKLALILHGNLQYEARTFENLDACEVLFS 349  
QY 351 TATGRSAILGLPTQULTALTORSMFLRGEELITPLNQOAVDSDSLAMALYACCFEN 410  
DB 350 PSLATASLELVNPPDLMSCLTSLRITLTGELVSTPLSRQALDVRDAFYKIGRLFW 409  
QY 411 VIKINSRI- - - - -KGNEDFK- - - - -SIGILDFGFENFEVNHFEQFNINAYANEKLOEYFNKH 464  
DB 410 IYDKINAIYKPPSOYDKNSRSIGILDFGFENFVANSFEQICINPANHLOQFVRVH 469  
QY 465 FSLSELYSREGVLWEDIMIDNGECULI- EKKGLIALINEESHPPQATDSTLLEKHA 523  
DB 470 FKLEOEYDLESIDWHLIEFTDQDALDMANKPMNIIISLIDESKFKPKTDITMHLKHN 529  
QY 524 SOHANNHFYKPRVAVNN- - - - -FGVRYAGEVOYDVGRGLEKXRPDPDOLLMLLESPR 579  
DB 530 SQHKLANNITPPK- - - - -NNHETQGINHFAIGIYVEETQGLEKRDITLHGDITQLVSSRN 586  
QY 580 DFIYDLEHVSHRNOPTLKCQSKRR- - - - -PTVSSQFVDSLSHSLMATSLSNFPVRCIK 637  
DB 587 KFIKQIF- - - - -QADVAMGAEIRKSPITLSSQF- RSLBELMRTILGCGQFFVRIC 637  
QY 638 PNNQKRPDQAVLNOLARYSGMLTIRIKAGYAVRPPDQFYRYKYKLMENLA- - - - -L 694  
DB 638 PNFKKKPMLEFDRHLCYRQOLRYSGMETIRIRAGYDIRYSFVEFVRVYLLTGLGVPAVK 697  
QY 695 PEVVRGKSTLLDLYASNSEMOLGKTVLRESLOKLEKREBEVSHAAMVIRAHVIG 754  
DB 698 QGDLRGTCORMAEAVLGTIDHMQIGKTKIFLKHDDMLLEVERDKAITDRVILLQKVRIG 757

QY 755 FLAKQYKRVLYCVIIQ- - - - -KNY- - - - - 775  
DB 758 FDRSNFLKKNAAATLIQHNHGRNCNKYGMRLGFLRLQALHSRKLHQYRLARQRI 817  
QY 776 - - - - -RAFLRRFLHLKAAIYFQKOLRQIARVYROLAK- - - - -REGEKKQOBE 826  
DB 818 IOFORCRAVYLRKAFRHLMAVLTVOAYARIMARRLHORARAYLWRLERLAE 877  
QY 827 EKKKE- - - - -EEREREREREAELPAQOEETRKOELALQKQEAELTRELEKO 880  
DB 878 ELKREMSAKKAKEERERHGERLQOLAREDE- - - - -RELKEMARKKELLQOMERA 932  
QY 881 K- - - - -ENKOVEIL- - - - -RLEKEIEDIORMKEQOELSTASLOKLOER 921  
DB 933 RHEPVNHSMDVMKMGFLTSGGLPGQEQAPSGFEDLERGRM- - - - -VEED 981  
QY 922 RQELRLEEBACRAQOEFLESINDELDECVRNIRLSGSGSESSSLAESACEKPNF 981  
DB 982 LPAALPLPED- - - - -EBDLSEYKFAKF- - - - -AATYQGTTHSYTRRP- - - - - 1021  
QY 982 NFSOP- - - - -YBEEVDE- - - - -GFEADDAFQDSPSPSEHSHDQRTSGIRTS 1025  
DB 1022 -LKQPLHYHDDGDLAALAVITTLRFNG- - - - -DLPEPKYH- - - - -TM 1061  
QY 1026 DDSSEDPYMDTVVPTSPSADSTVLA- - - - -PSVQDSGLHNS- - - - - 1065  
DB 1062 SDGSEKIPVMTKIYETLGKTKYRELQALQGEBAQLPEGQKSSVRRHLVHLTLKKSK 1121  
QY 1066 - - - - -SGSEGYCNPQNGDLPSPD- - - - -GDYDQDDYEDDA 1099  
DB 1122 LEEVYTKRLHDESTVQNGSMLEDRPISLXELHPIIGNGILRPALREIYQISKO- - - - - 1178  
QY 1100 ITSGSVTFPSNYSYGOWSPDYRCVSGTYNS- - - - -GAYRPSSEGAQSS 1143  
DB 1179 LTHNS- - - - -KSYAAGMILVSLC- VGCRAPSKPYKTLNFIHGCPGAPCEBRLKRT 1234  
QY 1235 FVNG- - - - -TRQPSWLELQATKSKPIMLP- - - - -VTFMDGTTLLTDSATKELCN 1285  
DB 1190 SKRKMVLYKD- - - - -ETPLMFPSKQELKQGLH- - - - -KKGGSSTLSRW 1233  
QY 1286 ALADK- ISKQFGLSYIALFDKVSILSGSDHWDALISOEOYAKEGA- - - - -OERNA 1340  
DB 1234 KRWFLRQSKLMYFENDESEELKGTV- - - - -EVRTAKE- - - - -IIDWTKENGID 1279  
QY 1341 PWRLF- FRKEVFTPMHSFSDNVATNLIYQVVRGVKGEYRCXEDDLAELASQOYFVD 1399  
DB 1280 I- - - - -IMADRTFHLIASPED- - - - -ASQMFVLSQVH- - - - -ASTDOEIQENHDEGA 1323  
QY 1400 YGSEMILERLNLVPTYIPDRREITPLKTEKQAQALIAHKKGIYAQRRTDAQYVEDV 1459  
DB 1324 NQONAVGLTDVGLIISVCASDSDPDRNSFVITITARVHLHCNADTDEENHMTTLQORSKG 1383  
QY 1460 SYARFKWPLFLSRFYEAYKFGSPSLPKQDVIVAVMTGVYFVDEBOV- - - - -LLELSRP 1514  
DB 1384 DTRVEGOEFTVRCMLHKEVKN- - - - -SPKMS- - - - -SLKLKRWFLVTH 1422  
QY 1515 EIMAVSSSEKCVWLSTLGSDDLGAAPHSAGMLTPAGBCSCWSCRAKKTAPBFTLAT 1574  
DB 1575 IKGDEYTFSSNAEDIRDLVTF- - - - -BGLRRSKYVALADNPAPAEESGFLSF 1627  
QY 1478 TK- - - - -LNEAL- - - - -RMSVYQNTVDTAIPDT- - - - -PTQOLI- - - - - 1509  
DB 1628 AAGDILIIDHDYGOVNSGMANGINERTKQGDPEPTOCVVMPTVTPPREIVALVMT 1687  
QY 1510 - - - - -ODI- - - - -KENCILNSDVVEQIYKRNPIIAR 1534  
DB 1688 PQRQDVVRLDQRLTAPEVRAKPYTLLEFSYDVRPPPKHLSKVMVMSKAGKQRLMSH 1747



QY 1535 THPHLSPLPLPYEDININ-----LLKDKGYTLQDEAIKIFNSLQOLESMDPI 1585  
 DB 1748 TREPPLKQALLKLLGSEELSOEACIAFIAYLKMVG-----DYPKSTRSVNELTDQ----- 1798  
 QY 1586 PIIQIILQGHDLRLPRLDELQYOLIKO--TNKVPHGSGVGNLYSQWILCLSCSTPLPSHG 1643  
 DB 1799 -IPEGGLK-----EPKADAVYQILQLTDNHRVSEERK---WELLWLCTGLPEPPSNI 1849  
 QY 1644 ILKYLKFKLRIRREOPFGTEMEKVALFTYESLKK---TKRREVPSPRDEIALIHR-QEM 1699  
 DB 1850 LLPHQRFLOS-RKHCP-----LAIDCQRLQKALNRSGRKPHLYVEVALQHTTQI 1902  
 QY 1700 TSTVYCHGGSCCKITINSHTTAGEVVEKILRGLAMEDSRIMPAL----- 1743  
 DB 1903 FPKYVFPDDTDPAFEVSESTKAKDFQNLATRLILLKSESG-FSLFKIADKYISVENDF 1961  
 QY 1744 -PEYNGHVDKAIESRRVADVLAKEPKLANATSEVGLPMKIFV--GLYGFELDTNMP-KD 1799  
 DB 1962 FDFPFRHLTDWIKKARPIDGI-----VPSLTVOVFPMKLDW---TTTVPGSD 2006  
 QY 1800 SV-EFAFMFEQAHNAVIGHHPAPBENLOVAL--RLQTLQGDYTLHAIPPLBEVSL 1856  
 DB 2007 PMADSFHYQELPKLRGHNCTREBVQLGLIRVKE--BEDKSYFSPISPL----- 2059  
 QY 1857 QRLKARISQSTKTFPPCEERLEKRRISFLEGLTRRSFRGAVVRQKVEEQMLDMWIKERV 1916  
 DB 2060 -----LR-----ELVQDLIRQVSPDDW----- 2077  
 QY 1917 SSABASIIDKMRKFGQMNQOAMAKYMLIKEMPGYSTLPDV-EKEGGFQOELMGLS 1975  
 DB 2078 ---KSTIVAYFNMHAKSKSEALAKLKIIFKMPFTFGSAFPEVKQTEENFPELILIAN 2134  
 QY 1976 ADAVSVYKRGSRPLEVFEYEHLLSPGAPLANTYKIV-----VDERELLFETSEVADVAK 2030  
 DB 2135 KGVGLSIDPKTDILTHPTKISNWS--GNTYFHITTNLVRSKLLCETSLGKMD 2192  
 QY 2031 LMKAYISMIVKKRYSTRSASSOGS 2055  
 DB 2193 LLTYSIQWL-----TAMSKORGS 2211

RESULT 5  
 M15\_MOUSE STANDARD; PRT; 3511 AA.  
 AC 09QZ24; 070395; 09QWL6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myosin XV (Unconventional myosin-15).  
 GN MYO15A OR MYO15.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_Taxid=10090;  
 RX MEDLINE=98267310; PubMed=9603735;  
 RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,  
 Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,  
 Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,  
 Friedell R.A.;  
 RT "Characterization of the human and mouse unconventional myosin XV  
 gene responsible for hereditary deafness DFNB3 and shaker 2.";  
 RL Genomics 61:243-258(1999).  
 RN 12;  
 RP SEQUENCE OF 1168-2970 FROM N.A., AND VARIANT SH2 TYR-1779.  
 RC TISSUE=embryo;  
 MEDLINE=98267310; PubMed=9603735;  
 RA Probst F.J., Friedell R.A., Raphael Y., Saunders T.L., Wang A.,  
 Liang Y., Morell R.J., Touchman J.W., Lyons R.H., Noben-Trauch K.,  
 Friedman T.B., Camper S.A.;  
 RT "Correction of deafness in shaker-2 mice by an unconventional myosin  
 in a BAC transgene.";

RL Science 280:1444-1447(1998).  
 RN (3)  
 RP SEQUENCE OF 1237-1823 FROM N.A., AND VARIANT SH2 TYR-1779.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=98369604; PubMed=9703981;  
 RA Wakabayashi Y., Takahashi Y., Kikkawa Y., Okano H., Mishima Y.,  
 Ushiki T., Yonekawa H., Komatsu R.;  
 RT "A novel type of myosin encoded by the mouse deafness gene shaker-2.";  
 RL Biochem. Biophys. Res. Commun. 248:655-659(1998).  
 CC -1 FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
 CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF  
 CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.  
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1 TISSUE SPECIFICITY: IN THE DEVELOPING INNER EAR, EXPRESSED IN  
 CC COCHLEA AND VESTIBULAR APPARATUS. EXPRESSION APPEARS TO BE  
 CC RESTRICTED TO COCHLEAR NEUROSENSORY CELLS AND UPPER EPITHELIAL  
 CC LAYER OF MACULA SACCUCLA. ALSO EXPRESSED IN MACULA UTRICULAR AND  
 CC CRISTAE AMPULLARIS OF THE SEMICIRCULAR CANALS. IN ADULT COCHLEAR  
 CC HAIR CELLS, HIGHEST EXPRESSION IN STEREOCILIA AND APICAL BODY.  
 CC -1 DISEASE: DEFECTS IN MYO15A ARE THE CAUSE OF SHAKER 2 (SH2), A  
 CC CONDITION CAUSING DEAFNESS, CIRCLING BEHAVIOR, HEADTOSING AND  
 CC HYPERACTIVITY. AUDITORY HAIR CELLS OF AFFECTED ANIMALS HAVE VERY  
 CC SHORT STEREOCILIA AND A LONG ACTIN-CONTAINING PROTRUSION AT THEIR  
 CC BASAL END.  
 CC -1 SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF144095; AAF05904.1; -;  
 DR EMBL, AF053130; AAC04024.1; -;  
 DR EMBL, AB014510; BAA36582.1; -;  
 DR HSSP, P10587; 1BR2.  
 DR MGD, MGI:1261811; MYO15.  
 DR InterPro: IPR000299; Band 4.1.  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR000857; MYTH4.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00612; IQ; 2.  
 DR Pfam: PF00784; MYTH4; 2.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART, SM00015; IQ; 2.  
 DR SMART, SM00242; MYSC; 1.  
 DR SMART, SM00139; MYTH4; 2.  
 DR SMART, SM00326; SH3; 1.  
 DR PROSITE, PS50096; IQ; 2.  
 DR PROSITE, PS50092; SH3; 1.  
 DR PROSITE, PS00660; BAND\_41\_1; FALSE NEG.  
 DR PROSITE, PS00661; BAND\_41\_2; FALSE NEG.  
 DR PROSITE, PS50057; BAND\_41\_3; FALSE NEG.  
 KW Myosin; ATP-binding; Actin-binding; coiled coil; Repeat; SH3 domain;  
 KW Disease mutation; Deafness.  
 FT DOMAIN 1 1871  
 FT DOMAIN 1872 2013 HEAD OR MOTOR DOMAIN.  
 FT DOMAIN 2014 3511 NECK OR REGULATORY DOMAIN.  
 FT DOMAIN 1307 1334 TAIL.  
 FT DOMAIN 1776 1783 COILED COIL (POTENTIAL).  
 FT DOMAIN 1886 1908 ACTIN-BINDING (POTENTIAL).  
 FT DOMAIN 1909 1938 IQ 1.  
 FT DOMAIN 1938 1938 IQ 2.

FT	DOMAIN	2848	2934	SH3
FT	NP_BIND	1289	1306	ATP (POTENTIAL).
FT	VARIANT	1779	1779	C -> Y (IN SH2).
FT	CONFLICT	1330	1331	MISSING (IN REF. 2).
FT	CONFLICT	1579	1579	L -> R (IN REF. 3).
FT	CONFLICT	1955	1972	MISSING (IN REF. 2).
FT	CONFLICT	2077	2077	L -> M (IN REF. 2).
FT	CONFLICT	2139	2139	V -> P (IN REF. 2).
FT	CONFLICT	2953	2953	L -> A (IN REF. 2).
SO	SEQUENCE	3511 AA;	395533 MM;	38C9622F98A2D395B CRC64;
Query Match 14.8%; Score 1593; DB 1; Length 3511.				
Best Local Similarity 24.8%; Pred. No. 3,1e-65;				
Matches 609; Conservative 373; Mismatches 836; Indels 642; Gaps 81;				
QY	59	EGVDNMASTLHSGSIMYNVLFQYKKNQITWTYIGSLIASVNPQPIAGLEPATMPOY	118	
DB	1205	EDGVEDMQLBQDGLQTTVLANKTRFERNLITYIGSLIVSNPRMFR-ITGPEQVOYQ	1265	
QY	119	SRRLGELPRHIFAIANECYRCIMKRNOCILIGESGAGCTESTKILKFLSVISQOS	178	
DB	1264	SGRALGENPRLFAIANLAFAMLDAKONOCYILIGESGSGCTEATKILRLCLAMQR	13222	
QY	179	LELSKENTSCVERALLESSPIEMAFGNKTYNNNSRFGFGVQNLNCOQNGIOGRIV	238	
DB	1333	-----DVMQOIKLELTPLEFNGAKTYNRNDSRFGFVBIPL--EGVIGCAT	13737	
QY	239	DCILSSORNVVRNGERNYHIFVYLLAGLEHEEREFEYLTSPENVHYLNOSGVEDTI	298	
DB	1374	SOYLLEKSRIVQAQNERNYHIFVLLAGLPRLQALQASLSQAEITYYLLNGGCEINAK	14333	
QY	289	SDQESFREYITAMDVQFSKEEVEVSRLLAGILHGNIEF---TAGGAQVSFKTAL	353	
DB	1434	SDADFRRLTAMEVLTGFTSEODSIFRLIASILHGNVYFEKHETDAQEVASVVSAREI	14931	
QY	354	GRSAELGLDPRQLDALTORSMFLARGEILTPPLVQAQVNRSDSLMALVAACSEWIK	413	
DB	1494	QAVALLQVSPGLQKAITFKTETIRKIFPLPLVESAVADRDIAVLVALFGWLIT	1553	
QY	414	KINRIKKNEDPFKSGIGLDFGFENFENVHFPQFPIANANEXLOEYFNKHI.FSLEQLEYS	473	
DB	1554	RVNALVSRKQDPLSLAIDLIFGEDLSNSPEQLCINANENLQYLPNKIYPOEQBEYI	16133	
QY	474	REGVWEDIWDINGECIDLIE-KKLGLALINBSHFPOATDSTLLEKLSOHNANHFY	532	
DB	1614	REQMDREIAFADNPICNLISLKEYGILRIIDDCCPQATDHTFLQKCHYHNGANFLY	1673	
QY	533	VKPRVAVNVFGVKNHAGEVOYDVRGILERNRTPFDLNLNLRERPRFIYDLFEHVSR	592	
DB	1674	SKPKPLREEFTIKHAYAGVLYQVNHFLDKNDHQVQDVLDFVHSRTVVALFESHAQ	17333	
QY	593	NNQDTLKCGSKS---HRRPYVSSQPKVSLSLYMTLSSNPFPRCIKPMNQKPRDQFD	648	
DB	1734	TAPPLGSGSSSTRLRYKATVAAKQ-OSLIDLVAKMERCNPLFRCLKPNHKEBGLFE	17929	
QY	649	QAVVNLQRYSGMLETVARIRKAGVAVRBPQOFYRKYVLYMR-NLALBEDVYGKCTSL-	706	
DB	1793	PDVMAQAQRYSGVLETVIRIKRGFPVRLPQVGFIDRYCVALVKLVNPAAD-GDMCVSLIS	1851	
QY	707	QLYDASNEMLQKTVLFRBSLEQLEKREKEEVSAAAMVIRAHVGLFARKQYRKVLY	766	
DB	1852	RLCTVTPPMYRYSKFLKEHLDLSMERVQNRALTLQRLRGFTIORHFRSLRR	19111	
QY	767	CVVILQKNYARVFLERRFLHKKALIVEOKQALRGQIARVYROLLAEKREO-EKKKQEE	825	
DB	1912	KIILQSGARGLFARQIQOMQOSLAKRSLVHTYVNRRYIKLAEQRRRQAEWMLBEQ	19717	
QY	826	EEKKKREBEEREREREREAEALRAQOEETTRQDELALQOSKEAEVLT---RELKEKE	882	
DB	1972	EELSKRELV-----PVNHLERPA-----EVALLQAAAGLKLKSSGRRVAVVAP	2015	
QY	883	NKQVEELIRLEKEIIDLQRMKEQDELSTVASLQKLERDDELRLE-EERCAAEFL	941	

Db	2016	RLQBPCTVLPJLDINNYP-MAKFRCHFKBESFGMLTVPLKMPJLRLPVEHHAALISVFK	2074
Oy	942	ESLNFDIIDEICVNRNIERSLSGSEHSESELASACEK-----PN-ENFSOPY-	987
Db	2075	LILRFMG-DPHLHGTOEWILGNYIVHQVLVEPALRDEILAQLANQVWRNPAAVYSKRGWL	2133
Oy	988	-----PEEVD-----GPAADDAFKOSPNSSEHSHSQRSQRTS	1025
Db	2134	LMAACSGFASPPLHDKFLKYSVDYQNGQA-----VCHRLLOAMGSGAART	2183
Oy	1026	DDSEEDPYMNDVTVPFSPADSVTLAASPQDSGLHNSSGESTYCMPONAGDLESPD	1085
Db	2184	FPPQO---LEWTAIOEKASMAADVSCFNGDQFSCPHNTSTGAIV---ADILKHR	2233
Oy	1086	GDYDYODDYEDGAITSGSVTFSNSYGSQWSP---DYRCSV-----GTYN	1128
Db	2234	G-----LADG---WRGWTVAMGN---GVQMAELAGHDYLDVLVDSLELLRLDFPROKSXF	2281
Oy	1129	SSGAFRRSSBEAOSFEDSEEDPFRSDRTDDELSTRDSVSCUTLYFHSFLYMK--GG	1186
Db	2282	IYGA-----EGPLRGRDTRGVFGNCWDSDEPTRPOROPDHVAKMPLDGYCSHKDGT	2336
Oy	1187	LMSMKRKM-----CYLKHETPLMFRSKQ	1210
Db	2337	NGETBAQWMSNROAVDSIGESTYPPRELDGYLDSLFDVPLACDADLEKTAIAYMK-	2395
Oy	1211	EALQGMJLHKKGSGST--LSRRNMK-----KRWFLRQSKLMYEND	1251
Db	2396	-----GGGQPGGGGSGSTSEDTSRPBPBKPIRGLDASTLALQOAFIHRQAVLAREMT	2450
Oy	1252	-----SEBKUKGT-----VEVR-----TAKEIIDNTTKEN-----	1276
Db	2451	LQALALQOQPLSATSRRPOLPERPLAPRKPVTWGTPRPAKPLVVRPTPOSWABGSYAKA	2510
Oy	1277	-----GIDIIWADRTF-HLIASEEDASQWFSVSOVN-ASTDOEIOEM-----	1318
Db	2511	PKIPSKPVAVPILQODWTAPBSIASBELVR--YSTLSEHFPOPTQOINSIYQYKOPR	2568
Oy	1319	-----HDE-----QANPONAUGTLDVGLDSV-----	1340
Db	2569	WAGHPEARRTDGGKVFRRPRPDHHEALMLIKGQTOGLAVVPCTQVSHREAVAMVAPVTSAP	2628
Oy	1341	-----CASDPRDRPNSV-----ITAN-----	1358
Db	2629	RPCMGPTPVOPSRSLPEPDPVQTOHLRLVNPVEYQODIPWRILFRKEVEFYPRKONYSHP	2688
Oy	1359	-----RVLN-----CNADPTEENHMTTLQSRKGPTRVBSGQFIVRGMLHKYKN	1404
Db	2689	VQDRLFRQILDHTFSEACRISDERLQMKALFAPQOBLDQO-----RPLTVESYKR	2740
Oy	1405	\$-----PKMSL-----KLKRWFLV-----	1420
Db	2741	AAISMARDSWETIYRSRLFPMAGSVGTGVOILAHSHTGILKQWYKSGSKASRRRLVLCAY	2800
Oy	1421	-----THNSLDYKSSSEK-----NALKJGLTV-----LNSLCSV	1450
Db	2801	SFADILFVTPMSQMLTEFNLSNEKTLIFSARAQOVKTLVDTFILKKSDSDYVAAVENFL	2860
Oy	1451	PPDKIKR-----ETGYMNTVYGRKICY-----	1474
Db	2861	SEDELLLSFHKGDIILHQLSEPTRVGVSAGCVARKKLVYLEELRRRGPDPFGMRFGAVHGR	2920
Oy	1475	--RLYTKLNEAT-----RMSVJONVTDKAPJDT-----PQOOLIOD	1511
Db	2921	VGRFPSELVOPRPAAPDLQPAEBGGRGAAAVAAVAAAAGVGRRRBEPFRASAD	2980
Oy	1512	IKENCL--NSDVE--QIYKRD-----ILRYTHRPLHSPL	1543
Db	2981	SGEDSIALPSTWLEFAPQKFRDPRRRPRDGLKLKSKEDEBSKTLBDVLCFTTVPIQESL	3040
Oy	1544	LPLPYGINLNLKDKGYTTLQDEBAKIFNSLOQ-----LEMSDPIPIIIOGLIOGH	1597

Db 3041 IEL--SDSLNKM-----AVDMFVAVRFGADAPLKGQSE-LDVLCTLLKLCSD 3086  
 Qy 1598 LRRLDELXCOLIKQTNKVPHPGSGVNLXSWQITLCTSCPLSPRSILKYLKHLKRIE 1657  
 Db 3087 HEVWDECCQIVKQITDSSSPQDSCQGMRLIYMAAYGSEVFYLLRFLQHV-S 3145  
 Qy 1658 QPFGTEMEKVALFYVESLKT--KCRFVPSHDEIEALI-HRQENTSTVYCHGGGSKI 1713  
 Db 3146 WPFGLPFGQIAKACEQNLQKTLRFGRLTFPSSMELRAMLAGRSSKROLSLPGGLERHL 3205  
 Qy 1714 TINSHTAGEVVEKILRGILAME-----DSRNPALFENYCHDKALIESRTVADVLAKRE 1768  
 Db 3206 KIKTCTVALDVIGELCTTEMLTRPEAFDEIVLIVTNRCGHV-CPLSCRAYILDVASEME 3264  
 Qy 1769 K-----LAATSEVGDLPKMFYFKLYCPLDTONVPDQSVFAFMFQAEVAVHGHHPAP 1822  
 Db 3265 QYDGGTLMFRVLMNQPLKFEHELIVTHNYGVLPDYLGKF-----SSVPAQPRFE 3317  
 Qy 1823 EENLOYLAALRLQYLOGDYTLAAAIPLPEEYVSLQRLKARISQSTYTFPCRLEKRRTS 1882  
 Db 3318 QOLQVSKLASIQHRAKD--HFYLSVREVQ--EYIPAOVLVHTTAG----- 3359  
 Qy 1883 FLEGTRRSFRGSGVVRKQVVEBQMDMFIKEVSSAPASIIDKMKFGGMNOEQAMAKY 1942  
 Db 3360 -----DTWL-NLVSGHR-----QOTQALSPHQARAP 3385  
 Qy 1943 MALIKEPGYSGLPDE-CKEGGPFQELMLGVSAVVS-----VYKR 1984  
 Db 3386 LGLLSFPPLFGSSFFFGSCSNLVVAPCLAVNHNGLMPLSKTIELVYKLPKLTQST 3445  
 Qy 1985 GGGPPLVEFYQYEH1-LSPGAPLANTYKIVDERELLFETSZVVDVAKMKAYISMVKKR 2043  
 Db 3446 WTKQFTNANSYFVEISLD-----VAAQRTWQLQJELGRVAVVAHVESLTSAR 3497

RESULT 6  
 MY15\_HUMAN STANDARD; PRT; 3530 AA.  
 ID MY15\_HUMAN STANDARD; PRT; 3530 AA.  
 AC 09UKN7;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myosin XV (Unconventional myosin-15).  
 GN MYO15A OR MYO15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.  
 RX MEDLINE=20021762; PubMed=10552926;  
 RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,  
 Barber T.D., Miller W., Touchman J.W., Jin L., Sullivan S.L.,  
 Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,  
 Friedell R.A.;  
 RA "Characterization of the human and mouse unconventional myosin XV  
 RT gene responsible for hereditary deafness DFNB3 and shaker 2.";  
 RL Genomics 61:243-258(1999).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A., AND VARIANTS DFNB3 TYR-2111 AND PHE-2113.  
 RP MEDLINE=98267311; PubMed=9603736;  
 RA Wang A., Liang Y., Friedell R.A., Probst F.J., Milcox E.R.,  
 RA Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,  
 RA Friedman T.B.;  
 RA "Association of unconventional myosin MYO15 mutations with human  
 RT nonsyndromic deafness DFNB3.";  
 RL Science 280:1447-1451(1998).  
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
 CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF  
 CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSOR HAIR CELLS.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED  
 CC AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,  
 CC PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE  
 CC PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.  
 CC -1- DISEASE: Defects in MYO15A are the cause of an autosomal recessive  
 CC form of nonsyndromic deafness (DFNB3).  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL, AF144094; AAF05903.1; -.  
 CC EMBL, AF051976; -; NOT\_ANNOTATED\_CDS.  
 CC HSSP, P10587; 1BR2.  
 CC Genew, HGNC:7594; MYO15A.  
 CC MIM, 600316; -.  
 CC InterPro: IPR000299; Band 4.1.  
 CC InterPro: IPR000048; IQ region.  
 CC InterPro: IPR000857; MYTH4.  
 CC InterPro: IPR001452; SH3.  
 CC InterPro: IPR001609; myosin head.  
 CC Pfam: PF00612; IQ; 3.  
 CC Pfam: PF00784; MYTH4; 2.  
 CC PRINTS, PR00193; MYOSINHEAVY.  
 CC ProDom, PD000355; myosin\_head; 1.  
 CC SMART, SM00015; IQ; 2.  
 CC SMART, SM00242; MYSC; 1.  
 CC SMART, SM00139; MYTH4; 2.  
 CC SMART, SM00326; SH3; 1.  
 CC PROSITE, PS00660; BAND\_41\_1; FALSE\_NEG.  
 CC PROSITE, PS00661; BAND\_41\_2; FALSE\_NEG.  
 CC PROSITE, PS50057; BAND\_41\_3; 1.  
 CC PROSITE, PS50096; IQ; 3.  
 CC DR PROSITE, PS50002; SH3; FALSE\_NEG.  
 CC Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;  
 CC Calmodulin-binding; Disease mutation; Deafness.  
 CC KX Calmodulin-binding; Disease mutation; Deafness.  
 CC KW KX  
 FT DOMAIN 1 1887  
 FT DOMAIN 1888 2029  
 FT DOMAIN 2030 3530  
 FT DOMAIN 1323 1350  
 FT DOMAIN 1792 1799  
 FT DOMAIN 1902 1924  
 FT DOMAIN 1925 1954  
 FT DOMAIN 1955 1976  
 FT DOMAIN 2867 2953  
 FT DOMAIN 3206 3443  
 FT NP\_BIND 1315 1322  
 FT NP\_BIND 2111 2111  
 FT VARIANT 2113 2113  
 FT VARIANT 2113 2113  
 FT SEQUENCE 3530 AA; 395171 MW; 3D103523D48CBE4A CRC64;  
 SO Query Match 14.8%; Score 1589.5; DB 1; Length 3530;  
 SO Best Local Similarity 25.1%; Pred. No. 4.6e-65;  
 SO Matches 623; Conservative 348; Mismatches 818; Indels 697; Gaps 88;  
 Qy 59 EGVGDMASTIEHSGSITMNVLPQYKRNQMTYIGSTLASVNPQPIAGLYEPATMEQY 118  
 Db 1221 EDGEVDMQLEQLQSTTVALSNKIFERNLLTYTGISLVSVPYQ-MEGTIGPEQVOQY 1279  
 Qy 119 SRRHIGELPPIHFAIANECYRCLMKRHDNQCILKIGESGAGKTBSTKLILKFLYSISQDS 178





Db 896 FRRUKAQAALKALTEARSABHLKRLNVGMENKVVQLOQRKIDQNKFKTLSEQLSAVTS 955  
Qy 867 --SOKEMALRELEKOKENQVEEILLEKEIEBLOKMKQOQBELSLTEALOQLQERDQ 924  
Db 956 THANEVEKLKELARYOQNOBADPSLOOEVSOLR--TELQKHSRRVLEAHNNENG 1013  
Qy 925 ELRR----LEBACRAAQEFLESINPDEICVNIERSLSGSEFSEFSAESACEKPN 980  
Db 1014 ELKRVAIDLHEHNLKDF-KEHNL---HQLRQSAEBSQSSVEBNLLIKELBEERS 1068  
Qy 981 --FNFQOPPE-BEVDGEFADDAFKOSPNSEHG--HSDQRTSGIRISDDSEDPYM 1035  
Db 1069 RYQNLVKEYSQLEBQRYENLRDEQOTPEGRKPKPSOSSLESDSVYPSISTEIDTEALQ 1128  
Qy 1036 NDTVVPSPSA-DSTVL-----APSVODSGSLH-----NSSGSECTYCPONAGDLPS 1083  
Db 1129 QVEEIGIEKAMDTVVLKLOKRVRELEQERKLOVLEKEQDDSKVQVEQNGEDVD 1188  
Qy 1084 PDGDYDQDDYEDGATSSSVTFNSYSQSSPDYRCVGTYNSSGAYRFSSEGAQSS 1143  
Db 1189 QADADIANSLKQGE--LESENKKLKNDLNERMK-----AVADQAMQDN 1229  
Qy 1144 PEDSEEDFSRF-----DTDELSYRDSVYSCVTLPEFHSFLYMKGLNMSKRWCVL 1198  
Db 1230 STHSPDSVSLNLNQLKLANEBLEVRKEEV-----LILRQIMNADQRLSGK 1277  
Qy 1199 KDETFL-----WFRSKQKALKQGMWHLKKGSSSTLSR-RWKKRWFLRQSKLMYFENDS 1252  
Db 1278 NMEPNARTSPWSEKHVDQEDAEIYHGVCOQNSQTEBWG-----YLNEDG 1325  
Qy 1253 E-----EKLKGTVEVTRAKEIIDNTTKENGIDIMADRTPHLIAESPEDASQWF----- 1301  
Db 1326 ELGLAAGLQVARNLEAQOAOQLKHEEVEHLKQO--VEANKKEBKQOQFCQTL 1383  
Qy 1302 SVLSQVHASTDQEIQENHDEQANPQNAVGLTD-----VGLDSVCASDS 1345  
Db 1384 SPAQVEFGVQOEISRLTNEHLDFKELVEKLEKKEKLLKQKQIYMKVODLAAQALQ 1443  
Qy 1346 PDRPNSVITITARVLHCNADTEBEMHMT-----LLORSQDPRVEQGEIYVGMW--HK 1400  
Db 1444 SDR-----RHHLTRQVTVQKREKD-----FGQMBLYRK 1472  
Qy 1401 E-----VKN-----SPKASS-----LKLKRWFLTNSLDYKSSSEKNAKLGTVLNS 1445  
Db 1473 EDEALLIRNLVTDLKQPMLSGTVPCLPAYLLWNCIRAD---YTNDLKHSLLSSTING 1529  
Qy 1446 LCSVPPEDEKIFKETGYWNVTVYGRKCYRLYTKLNEATRWSSVIONVDTYAPIDTPT 1505  
Db 1530 IKRYVLKKNHEDFEWTSFMLSNTCRLHLCLKQYSGDEGFMQO--WTAKQN----- 1576  
Qy 1506 QQLIQIDIKENCJLNSDVVEQIYKRNPLRATYHHRHSLPLRPLRYGVDINLMLDKGYTTL 1564  
Db 1577 -----BHCNKMFDLTE--YFO-----VL 1592  
Qy 1565 QDEAIKIFNSLOOL-ESMSDPPI-----IGLILQTHDLRP----- 1600  
Db 1593 SDSLQIQVQQLIKIABGLQPMIVSAMLBNESIQGLSGVAPTYGRKRSSSMVDGENSYCL 1652  
Qy 1601 -----LRDELVCQLIKQTKNVPHPGSVGNLYSQWILTCLSCSTPLPSR 1642  
Db 1653 EAIIRQWNPFTVLCDQGLDPEIILQVFKOLFPMINAVTLNLD-----LRLKDCASW--ST 1706  
Qy 1643 GILKYLKFKHAKRIEOPFGTEMERKY-ALFTYESL-----KTKCKEFPSPSDEIE 1691  
Db 1707 GM--QLKYNISQLEEWLRGNLQOOSGAVQMEPLQAAQLQKCKKT-----QEDAE 1756  
Qy 1692 AL-----IHRQEMTSIVYCH--GGGSKITIN-SHTTAGVEVEKLIRGLAMEDSRNP 1741  
Db 1757 AICSLCSTSLSTQQLVKILNLYTLPLNGEERAVTYSFRTIOAQOQESNDPQQLLDSKHF 1816  
Qy 1742 -ALPEYN 1747

Db 1817 PVLFPFN 1823  
RESULT 8  
MS2\_DICDI  
ID MS2\_DICDI STANDARD; PRT; 2116 AA.  
AC P08799;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin II heavy chain, non muscle.  
GN MHCa.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxId=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87092286; PubMed=3540939;  
RA Warrick H.M., de Lozanne A., Leitwand L.A., Spudich J.A.;  
RT "Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).  
RN [2]  
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.  
RC STRAIN=AX2;  
RX MEDLINE=90353583; PubMed=2387408;  
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J., Gerisch G.;  
RT "Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II.";  
RL FEBS Lett. 269:239-243(1990).  
RN [3]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE=88112226; PubMed=2828113;  
RA Wagle G., Noegel A., Scheel J., Gerisch G.;  
RT "Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain.";  
RL FEBS Lett. 227:71-75(1988).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.  
RX MEDLINE=95345066; PubMed=7619795;  
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutok K., Holden H.M., Rayment I.;  
RT "X-ray structures of the myosin motor domain of Dictyostelium discoidium complexed with MgADP.BeFx and MgADP.AlF<sub>4</sub><sup>-</sup>.";  
RL Biochemistry 34:8960-8972(1995).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.  
RX MEDLINE=95345067; PubMed=7619796;  
RA Smith C.A., Rayment I.;  
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of Dictyostelium discoidium myosin to 2.7-A resolution.";  
RL Biochemistry 34:8973-8981(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.  
RX MEDLINE=96206189; PubMed=8611530;  
RA Smith C.A., Rayment I.;  
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";  
RL Biochemistry 35:5404-5417(1996).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.  
RX MEDLINE=97452580; PubMed=9305951;  
RA Gullick A.M., Bauer C.B., Thoden U.B., Rayment I.;  
RT "X-ray structures of the MgADP, MgATPgammAs, and MgAMPNP complexes of the Dictyostelium discoidium myosin motor domain.";  
RL Biochemistry 36:11619-11626(1997).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.  
RX MEDLINE=98070605; PubMed=9405148;  
RA Bauer C.B., Kuhlman P.A., Bagnshaw C.R., Rayment I.;  
RT "X-ray crystal structure and solution fluorescence characterization









QY 130 IFAIANECYCLMKRHDNOCILIKGSSGAKTEBTLILKFLSVISQOSELSEKTEC 189  
 Db 215 VALALNAVAYTMLKRVNOCIVYPGSSGKTOSTFLIK:CLTALSO-----KCYASG 267  
 QY 190 VERRAILESSPIMEAFNAKTIVNNNSRGKPFQULNICOKGNIOGRIVDCLTSSONRV 249  
 Db 268 VERTIIIGACPVELEAFGNNAKTAHNNSSRGKPIQVSYLSEGIYRGA--VEKILLEXSRIV 326  
 QY 250 RONGHRYNHYIFAYLAGLEHEBEREFYLTSPENHYLNGSGCEDKTSIDOE---SPR 305  
 Db 327 SOEKORNYHVFYLLLGVSSEBERQFOLKOPEDYTLNG---NHLKIEDGDLKHDE 382  
 QY 306 EYITANDVNOFSKEEVEVSRILAGILHGNIEFTAG-----GAOVSKTALGRSAELL 360  
 Db 383 RIKQANEMVGLPATKQIFAVYLSAILYGNVYKRRATGRBEGLEVPPEVLDLTSQLL 442  
 QY 361 GLDPQDTALQORSMFLRGEELTPLNQAANDSDSLMALYACCFPMVYKKTISRIK 420  
 Db 443 KVRREIIVELVTIKRTVYNDKILIPYSSEAITADSMKSLYSALFPMIVIRINHALL 502  
 QY 421 GNEDPK-----SIGILDIFFGFENFVNHPFQFNINYANEKLOEYFNKHI:PSLEQLEGR 474  
 Db 503 NKXDVENAVSCISIGVLIFFGFEDFERNSEFOFCIYANEOLOYNYNQHIFKLEQSEYOG 562  
 QY 475 EGLWEDIDWIDNGECLDLIEKK-LGLLALINEESHFPOATDSTILLEKLSQHANNHYV 533  
 Db 563 EGIYTHINIGYTDVNGCIIHLSKPKTGLFYLLDEESNFPHATSOITLAKFQOJEDHKYFL 622  
 QY 534 KRVANMNGVGVYAGVOYDVRGILEKRDTRDMLNLRSRDFTYDLF----- 586  
 Db 623 GTPVNEPAPFIQHFAGKVKYQIKDFREKMDYMRPDI:VALLRGSDSSVYRELIGMDPVAN 682  
 QY 587 -----BH----- 588  
 Db 683 FRNAVILRAIRAMAVLEAGRLAERAEKAGSSGQASHPELPRGASTSEKLYRDL 742  
 QY 589 ----- 588  
 Db 743 HNQIMIKIGLPMWGEDPRSLQSLRLOKPRAFILSKGIKQOII:PKMLDSKSLKLI 802  
 QY 589 -----VSRNNQDTLKCGSKHRRPTVSSQPKVDSIHLMTLSSNFFPRCTIKPNQKXP 644  
 Db 803 ISMTLDRITKSLHLHKKKKRPPSISAQFOT-SLNTLLEALGAEPPFRCTIRSNAEKKE 861  
 QY 645 DQPDQAVINOLRYSQGMLETVRIRKAGYAVRPFODYKRYKYLMMNLALPEV---RSG 701  
 Db 862 LCFDEDLVJQOLRYTGMLETVRIRSGYSAKTYFQDTEQFQVL-----LPKDAQCREV 916  
 QY 702 CTSLLQLYDASNSEWQGTKYVFLRESLQLEKREBEVSHAAVIRAVLGLFARKQY 761  
 Db 917 ISTLEKMKIDKKNVYIGKTKVFLKETERQALQETLHREVVRKILLLQSMFVRVLERHF 976  
 QY 762 RXYLYCVIIOKNYRPFLLRRRFLHLKKAIVFQKLRQOIAIRVTR-----QLLA 812  
 Db 977 LQMKRAAVTIOACWRSYVRRA-LERTQAAVYLQAAWREGWKLYRHQKOSIIRLOJSC 1035  
 QY 813 EKRQOEKKQOEKKEEKKREBERER-ERERREAEILRAOEBETTRQOELALQSKXA 871  
 Db 1036 RGHLOKFSQMSISEKQAKBEKRELELARAAGEGGQDAAAGQOABQGEPEBDG 1095  
 QY 872 BLTRELEKQKQKQVEILRLKEIEDLQRMKEQOELSTLEASLQLOERDQELARLEE 931  
 Db 1096 HLASBEVQPSDRSPLEHSSPEKEAVSPEKTLRPOQTVAASHHEKYPSSREKRESNR 1152  
 QY 932 EACRAAOEFLESANFD--EIDECVNRINISLSGGS--EFSSELAEKCEKXPNFNSQYP 988  
 Db 1153 -----QRLGHNKFKONKHIQSCKE--ESALRPSRRVYQBOGVSLLEBK-----K 1195  
 QY 989 EEEVEDECF-----EADDDAFK---DSPNPEHGHSDORTS:ZITSDSSSEDPYMDTVV 1040  
 Db 1196 ESEMEDTLVLEAENTSOQKTEQROMAVKUSSEETB--KTLPSGS---PRQOLER 1250  
 QY 1041 PTPSPADSTVLLAPSVQDSGLHNSSSGSESTYCMPONAGDLP--SPDGDYDQDDYEDG- 1098

Db 1251 PTLSLDNRV-----SPAPGASAPETPEDEKSKRCGSPRVOEKDSDGSGTQIOR--YLDAB 1304  
 QY 1099 -----ATSSGSYFSSYSGOWSPDYRCISVGTNSGASVAFSSRGASPEDSE---- 1148  
 Db 1305 RLASAVELMRGKULVAASPSAMLSQSLDLS--DRIRATGALLTPBERTISSTSDVSL 1363  
 QY 1149 -----EDPDSREDTDELRYRDSVYSCVTLPEFHSFLYMKGLMNSWRRWCVLKDET 1203  
 Db 1364 LPSLAKQAPAAETTGESAKKPAVOK----- 1390  
 QY 1204 LMFRSKQELAKOGMLHKKGGSGSTY-----SRNWKKEWFLRQSKLMTFEND 1251  
 Db 1391 -----KKRGDASSILPDAGLSPGSOVDSKSTPFKR--LFLHKTK-----D 1426  
 QY 1252 SEEXLGTVEVYTAKE-----IIDNTKENGIDIMADRPHILAE-----SPEDASQWFSVL 1304  
 Db 1427 KKYSLGABELLENASGVNVLBATIMKGLLEPSSQQRHAAAGEKRTKEPGGKKKRRNV 1486  
 QY 1305 SOVHASTDOELQEMHDEQANPONAVGTLDVGLIDVSCASDSPDRPNSFVITANFVLGN 1364  
 Db 1487 KIGKITVBEKMRRESVFRQITANLKLIDELFLKINLDRSQKPTIESLFIATSEKPSN 1546  
 QY 1365 ADT-----PEBMHHITLL-----QSKGDTREVGQEF-----YRGMLHK 1400  
 Db 1547 IKTMYSVNGKIHGVKXLMENYQIVVGNLATERGQKDTNVLNLFQSLDEFTRGYTKN 1606  
 QY 1401 E--VYNSPKNSSLLKXKRWFLTNSS---LDYSSSEKMLK-----LGTVLNLSLC 1447  
 Db 1607 DEEPVKQS-KQKKRKRQERAAVEHNGVAFSVOYSIQSCQCLSYIMLMDKALLCVC 1665  
 QY 1448 SVVPDEKIPKETYGMANTVYGRKHCYRLYTGLLEATRWSSVIONVTDKAPIDTPQO 1507  
 Db 1666 KTMCHKCVHKIQSHCSYTL-YORK-----GEPGAEHGFVCVDSLSDKASVETVLEK 1718  
 QY 1508 LIQDIKENCLSNVVEQIYKEN-----PIIRYTHHPLHS----- 1541  
 Db 1719 LLEHVEHMLG---YTEGLYRKSAGANRTRELQALQTDPAVKLENPFHAIITGVAKQWL 1775  
 QY 1542 -----PLRPL-PYGDINLNLKDKGYTTLQDEBAIKIFINSIQLESMSDPIPIQGILOTG 1595  
 Db 1776 RELPEPLMTFAOYGFPL-----RAVLEPEKQEDLAI---YALVLEHLEPNX 1818  
 QY 1596 HDL--RPLRDELVCQILKQTNKVPHPGSGVNLYSQIOLTC--LSCTPLPSRGLLYLKFH 1651  
 Db 1819 HNSLERLLFHLVKVALLDVAIRM--SPGALAIIFAPCLARCPNDSPLSMKOVAKITTCV 1877  
 QY 1652 LKRIREQPPTGEMEKYALFTYESLKKTKCREFPSPR 1687  
 Db 1878 EWLIIKEQ-----MRKYKV-KMBEISQLEMAESIARF 1907  
 RESULT 10  
 MYSJ DICDI  
 AC P54697;  
 ID MYSJ DICDI STANDARD; PRT; 2245 AA.  
 DT 01-OCT-1996 (rel. 34, Created)  
 DT 01-OCT-1996 (rel. 34, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)  
 DE Myosin IV heavy chain.  
 GN MYOJ.  
 OS Dictyostelium discoideum (slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_Taxid:44689;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX3;  
 RX MEDLINE=96215148; PubMed=8636147;  
 RT Hammer J.A. III, Jung G.;  
 RT "The sequence of the dictyostelium myo J heavy chain gene predicts a  
 RT novel, dimeric, unconventional myosin with a heavy chain molecular  
 RT mass of 258 kDa."  
 RL J. Biol. Chem. 271:7120-7127 (1996).

[2]  
RN SEQUENCE OF 1-1021 FROM N.A.  
RP MEDLINE=97039016; PubMed=884597;  
RA Peterson M.D., Uricoste A.S., Titus M.A.;  
RT Dictyostelium discoideum myoI: a member of a broadly defined myosin  
V class or a class XI unconventional myosin?  
RL J. Muscle Res. Cell Motil. 17:411-424(1996).  
RN  
RP SEQUENCE OF 182-298 FROM N.A.  
RX MEDLINE=95023928; PubMed=7937787;  
RA Titus M.A., Kuropa A., Loomis W.F.;  
RT "Discovery of myosin genes by physical mapping in Dictyostelium";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).  
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE  
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U42409; AAA65186.1; -;  
DR EMBL: L35322; AAA79858.1; -;  
DR HSP: P08799; 1MND.  
DR DictyDb: D001095; myoI.  
DR InterPro: IPR002710; DIL.  
DR InterPro: IPR000048; IQ region.  
DR InterPro: IPR004009; Myosin\_N.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00063; myosin\_head; 2.  
DR Pfam: PF00612; IQ; 6.  
DR Pfam: PF01843; DIL; 1.  
DR Pfam: PF02736; Myosin\_N; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosin\_head; 2.  
DR ProDom: PD003376; DIL; 1.  
DR SMART: SM00015; IQ; 3.  
DR SMART: SM00242; MYSC; 1.  
DR PROSITE: PS50096; IQ; 3.  
KW Myosin, ATP-binding, Actin-binding, Calmodulin-binding; Repeat;  
KW Coiled coil.  
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.  
FT DOMAIN 824 851 IQ 1.  
FT DOMAIN 872 901 IQ 2.  
FT DOMAIN 943 972 IQ 3.  
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).  
FT DOMAIN 1813 2245 TAIL.  
FT NP\_BIND 174 181 ATP (POTENTIAL).  
FT DOMAIN 669 749 ACTIN-BINDING.  
FT CONFILCT 191 191 L -> F (IN REF. 2).  
FT CONFILCT 284 284 A -> T (IN REF. 2).  
FT CONFILCT 291 291 G -> R (IN REF. 2).  
FT CONFILCT 332 347 NKGCFEIGVSDSEH -> IEMFELKVRMKS (IN  
FT REF. 2).  
FT CONFILCT 550 550 N -> K (IN REF. 2).  
FT CONFILCT 865 866 HH -> QQ (IN REF. 2).  
SQ SEQUENCE 2245 AA; 258478 MW; 615E5EF1D1AB55B CRC64;  
Query Match 13.9%; Score 1490.5; DB 1; Length 2245;  
Best Local Similarity 24.5%; Pred. No. 8.8e-61;  
Matches 539; Conservative 377; Mismatches 665; Indels 589; Gaps 71;  
QY 4 OGTRWMLREN-----GQHPSTVNSCAEGIVFRTYGVGFYTKQSTTHQKXTAMHPN 58  
DB 26 EGAGVWIPDELGMIGADVIEHSETSDADVLRTEQREV-----KIPLSKVFOKQNPDI 79

QY 59 EEGVDDMASLTELHGSGIMYNLFRYKRNQIWTYIGSILASVNPQIAGLYEPATMEQY 118  
DB 80 LGGVDDLSPLSHLBPAIHLNHHRYNLNQIYTIKILIAINPTSLP-LYKEMISAY 138  
QY 119 SRRLGELPPLHFAIANECYRCLMKRHD--NCCILKESGSGKTESTLLIKFSLVISO 176  
DB 139 YKQIGTTLAPHYAVALAEDAFKDM--RYDGTQSILVSGSGAKETTKYLLQYLAAMGN 196  
QY 177 QSLLESLKEKTS-----CVERALLESSPIMEAFGN 206  
DB 197 M-----IKSTSSSSINGINTSSDDGIPTPPSPMKKSQVDSVERVLESTPLEAFGN 251  
QY 207 AKTVNNNSNRFGKQVQNLICQKNIQGRIVDCILSSQNRVVRQNGERNHYIFALIA 266  
DB 252 AKTLRNDNSRFGKFIHFNMGSIIIGAKILAYLL-EGSIVRQVYERNHYIFQYLLS 310  
QY 267 GLEHERREPFYLTENNYHYNLQSGCQVEKTSIDQSPREVTANDVQFSQEEVREVR 326  
DB 311 GASEELKEKLNKLTIEEYSLNKSCGFIEGVSDSEHFNKTCMAQVAGITLVEQENVR 370  
QY 327 LLAGLILHGINIEFTAGGA-----QVSPKALGRSAELLG-LDPTQLTALQRSWFLRG 380  
DB 371 ILSAILLINFEPENIAGSNDSCQILDRPLEKYSVLGCAQDPBELNSMLTRKYVTCK 430  
QY 381 EEILPPLVQAQVDSRDSIAMLVYACCEBWTIKINSRIKGNEDKS--IGILDFGE 437  
DB 431 ESYISHTKERAENNRDLSIMFLYGMFDMVLVKINSMSISTQKSKSFIGVLDIYGE 490  
QY 438 NFEVNHFEQFNINYNANEKIQEYFNKGIPLSELEYSRELVVEDIDMIDNGCCLDIKK 497  
DB 491 SEFVNGFEQFCINYNNEKIQQLFNQHVFEQOEYIKEDIMSYIDFNQDTLDIENK 550  
QY 498 -LGLTALINBSHPQATDSTLLEKLSOHANNHYVVRVANNFGVGVAGEVOYDVR 556  
DB 551 PICILTLDEITMFKAITQTLATKLXGMSHSKFKRPFSTAFTHNHVAGKTYEYD 610  
QY 557 GILEKXNDTFRDNLNLRSERFPIYDLFEH-----VSSRNNQDTLK 599  
DB 611 QFLDKNKDPIIEQISLIORSNFSFIKVLMSHSDKFTQSPGHPQONGGPTSSNTKGT-S 669  
QY 600 CGSKRRPPIVSSQFVDSILHSLMATLSSNPFVVCIRPMQKMDQFQDAVVLQRLRS 659  
DB 670 GSSSMKFLSVSGQFST-SLATLMKTIITTPHYVCIKENPKLPQTEFKQOVHIQLRGG 728  
QY 660 GMLQTVRIKAGVAVRPFQDFYKRYKVL-----RNLLPBDVRGKSTL 705  
DB 729 GWBESVRIICAGFPFRRRLISEFYQRYKILYVDINTGSGGKSGSNKKIKRPLKVLQNL 788  
QY 706 LQLYDANSSEWQIGTKVFLRESLEQLEKREBEVSHAUVIRAHVIGFLARKOYRY- 764  
DB 789 LIGIELSDDKYKIGLTKVFLRAGQASLEDMRLQEDSATAVIOKRMKGVLRYKRYKQLR 848.  
QY 765 -----LYCVIITQKYNRA-----LLRR 782  
DB 849 DASLIITQTKRSVHAKHMLLSALQRTHSALLIQKVRFAHRDRVOYOKIRDSALQLOTWMR 908  
QY 783 RFL-----HKKAIVFO-----KOLRG-----QIARVYRQLLA 812  
DB 909 HLFSEQVHREKENNAIITQIKIROLISKREVDKLRGILILQAMRMKLARGVITQLRA 968  
QY 813 EKREBEKKQDEBEKKRKEBERERERERERERERERERERERERERERERERERER 871  
DB 969 EAR-----SLRTVQEQNKUQKLEELQWMLTSEA 998  
QY 872 ELTRELEKQK-----ENKQVE-EILRLKEKLEIDLQRMEOQELSTESLQVL 918  
DB 999 KKKQQLQEDQKVSDDTTISELSNNNDHLEIQLSQELQSDNSQSSQLQSEC-LSLTL 1057  
QY 919 QERDQ-----ELRLBEPACRAAQEFSLNPFDEIDECVRNIEISLGSSEFSESLAES 973  
DB 1058 EHQTLQDLHSSKLANKLEKDLSDQHSIEKLO-SQPNTEIQQLQFQKQSELSKLST 1116  
QY 974 ACEKRNPNFNSQPYDEEVDGFEADDDAFKDSPNPSEHGSDQRTSGIRTSDDSEBDP 1033

```

Db      1117 T--QOQIDFN-----KQEPD-----RUSQERDNTN----- 1140
Qy      1034 YWMDTVVPSPSADSTVLLAPSVQDSGLHNSSGESTCYQPNAGDPLSPGDYDYPD 1093
Db      1141 --NQLFIQQLKKAQNSTL----- 1157
Qy      1094 DYEDGATISGSSVTFPSNSYGSQWSPDRCSVGTYNSSGAFRPSSEGAQSPEDSEDPDS 1153
Db      1158 DY-----PISGIRNLE--RQVLEL 1176
Qy      1154 RPDTDDELSTYRDSVYSCVTLTYFHSFLYMKGGLNNSWFRNCVLDKDETFMLFRSQEAL 1213
Db      1177 R-DENQLIKERLDSIQ-----OSSQFQSGAALKEQOLQIV-QEJSEQLIKSEKEL 1227
Qy      1214 KQGMILHKGGSSSTLSRRMKRWPFRLRQSKLMPENDSEKYLKGVN----- 1261
Db      1228 GSEBEAKKOINLELELTDHKSQLOI--QOLTEGSENEKIKLKLKLESTYQDEKQOLOE 1285
Qy      1262 --VTAKEIIDNTTKENGIDIIIMDRTH--LIAEPETASQWPSVLQVHASTDOEIO 1316
Db      1266 LERIKSQKQSVED--EKNSLITQLTIVKPESTQVSTNVSHQEKITLTKSTBELNKSIG 1343
Qy      1317 EKHDEQANFQNAVGLDVLID-----SVCSADSPDRPNSVVIITANRVLHC 1363
Db      1344 KLOAQCKNDDEIRIKIQFELNDOQKQFTRQTKFSPDLOSCOSIDREKSEITI-----HS 1397
Qy      1364 NADTEEMHMTLLORSKGDTRVEGEF---IVRGMLHRYVNSRQMSLKKRWFL 1420
Db      1398 LERTMETLKSPERVOQSLKQERDQOQKDTINR--LENEYK-----QLTQKEF--- 1447
Qy      1442 THNSLDYKSEKRNALKGLTLVNLSCSVPPDEKIFKETGYWNTVYGRKHCYRLYKL 1480
Db      1448 ---EHEFFVAKQNSNQ-----TQESYULK-----EVTTCQOQNSREREL 1486
Qy      1481 LNEATRWSSVIONVDTKAPIDTPTOQLIQDKENCLNSDVVEQIYKRNPIIRYTHHPLH 1540
Db      1487 BEKKOHITRIDDERBELKKQLTQLOQHQSSQTQLALQNELERLKKE--LKY----- 1538
Qy      1541 SPLPLPYVDIMNLKDKGYT-----IQDEAKIFNSLQOLQESDPIPIIGIL 1592
Db      1539 -----KENGHETSQOQDPNMEIQSLRTTN--DQLKSLQD----- 1572
Qy      1593 QTGHDRLPLRDELVC--QLIKOTNKVPHBSVGNLTSWOILCTLCTPLPSRGILKYLKF 1650
Db      1573 -YEQEKKKLKDLSKQEAQORE-----SIIRMAELSAIQ 1610
Qy      1651 HLKRIREFPGTEM-----EKYALFTYESIKTKCREPVPSPDEIF-ALIHQEMTS 1701
Db      1611 HQQWENSEF--TDMQORNOELIESSALYQOLQOSTIDSTIKENEISKLQOOLETS 1668
Qy      1702 TYVCHGGSCKITINSHTAGVEVLEKINGLAMEDSRMNFALFEVNGHDXKIAESTVVA 1761
Db      1669 NOOLH---OLKEELNSMKOSNOE-----STBOSKQNLQLOEQOLK----- 1708
Qy      1762 DVLAKFEKLAATSEVQDLPMKFFYKLYCFLTDVNPKDSVEFAFMFEQAHAVIHGHDA 1821
Db      1709 -----SVTNEIS-----KQDDAVF----- 1723
Qy      1822 PBNLOVLAALRLQVQO-----DYTLHAIPLEEVGLJRLKARISOSTYTPPCERL 1876
Db      1724 --ENQKINNTIKEOETKRMVLELQOHIDEGKQOIQOLQJSTIAQLKQOQOSET--DRL 1779
Qy      1877 EKRRTSFLBGLTRRSRFTGSVVRQKVEBOMLDMWIKI---EYSSARASITII---KMKR 1929
Db      1780 EKEIQO-----MKERET-----OMKLVSTKLVYHMLERDMELYRNVEIYDKETEMK 1830

```

```

RESULT 11
MY9B_RAT STANDARD; PRT; 1980 AA.
AC 06358;
DT 16-OCT-2001 (rel. 40, Created)

```

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IXb (Unconventional myosin-9b).
GN MYO9B OR MYR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
RX MEDLINE=9518874; PubMed=7882973;
RA Reihard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
RA Baehler M.;
RT "A novel type of myosin implicated in signalling by rho family
RT GTPases.";
RL EMBL J. 14:697-704(1995).
CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC -!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
CC PERINUCLEAR REGION (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
CC LIVER, AND SPLEEN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 IQ DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, X77609; CAA54700.1; -.
CC HSSP; P08799; 1MDN.
DR InterPro; IPR002219; DAG_pe-bind.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR000159; RA_domain.
DR InterPro; IPR000198; RhogAP.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head_2.
DR Pfam; PF00130; DAG_pe-bind; 1.
DR Pfam; PF00612; IQ_4.
DR Pfam; PF00620; RhogAP; 1.
DR Pfam; PF00788; RA; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 2.
DR SMART; SM00109; C1; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00324; RhogAP; 1.
DR PROSITE; PS00479; DAG_pe_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_pe_BIND_DOM_2; 1.
DR PROSITE; PS50096; IQ_3.
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
KW Zinc.
FT DOMAIN 1 940 HEAD OR MOTOR DOMAIN.
FT DOMAIN 941 1045 NECK OR REGULATOR DOMAIN.
FT DOMAIN 1046 1980 TAIL.
FT DOMAIN 1841 1861 COILED COIL (POTENTIAL).
FT DOMAIN 1918 1948 COILED COIL (POTENTIAL).

```

FT DOMAIN 845 856 ACTIN-BINDING.  
FT DOMAIN 978 978 IO 1.  
FT DOMAIN 981 1001 IO 2.  
FT DOMAIN 1002 1024 IO 3.  
FT DOMAIN 1025 1054 IO 4.  
FT DOMAIN 1593 1641 PHORBOL-ESTER AND DAG BINDING.  
FT DOMAIN 1673 1822 RHO-GAP.  
NP BIND 239 246 ATP (POTENTIAL).  
SEQUENCE 1980 AA; 225035 MW; D79FEC4D0FAE0C05 CRC64;  
Query Match 13.7%; Score 1471; DB 1; Length 1980;  
Best Local Similarity 25.3%; Pred. No. 5,8e-60;  
Matches 497; Conservative 309; Mismatches 670; Indels 492; Gaps 60;  
QY 11 RENGHPSTVNSCAEIGVFRFDYGVFTYKOSTITTHOKV-TAMHPTEEGVDMASTL 69  
DB 101 REGGYFLDOERN-ADGSI-----QYLHVQLLAQPTACRLVGGGLPRPADDDLCNLP 155  
QY 70 ELHGSGIMVNLFORYKRNQIMTYIGSILASVNPYQPLAGLYEPATWEOYSRRHLGELPPH 129  
DB 156 ELNEANMLQSLKLPVQOKIYTAGSILVAINFEK-FLPIYNPKYVMENQOLGKLEPH 214  
QY 130 IPIANECRCILKMRDNOICILKSGSGAKTSTKILKFLSVIGQSLSELSKERTSC 189  
DB 215 VFPLADAVAYTMLRKVNOCIVISGSGSGKTOSTNFLHCLTALSO-----KGYASG 267  
QY 190 VERALIESPIMEAFGNAKTVYNNNSRFGKVFQVNLICQKNTIGGRIVDCILSONRVY 249  
DB 268 VERTIIAGFVLEAFGNAKTAHNNNSRFGKIQVNTYLESIGYRGA-VKEKYLLEKSRIV 326  
QY 250 RQNGERNYHIFYALLAGLEHEBEREPYLTSPENHYLNGSGVEDKTIISDOE----SFR 305  
DB 327 SQEKDERNYHVFYLLGVSEBEROEFOLKPODYFLNQ-----HNLIJEDGEDLKHDFE 382  
QY 306 EVTTANDWQFSBEVREVSRLLAGLHGNIEFTTAG-----GAQVSEPTALGRASLL 360  
DB 383 RLQQAEMWGFRLPATKKQIFSVLSAILYGNVYKKRATGRDEGLVEGPEVLDLISQL 442  
QY 361 GLDPTOLDLATORSMFLRGEILITPLNQOAVDSRLMALYACFEVVIKINSRIK 420  
DB 443 KYARELVEVLTGRKITTVDKILIPYSLEAITADSMAKSLYSALFDMIVIRINHAL 502  
QY 421 GNEDFK-----SIGLIDIFGFENFEVNEOFNINANEKLOEYFNKHFISLEQLEYSR 474  
DB 503 NKMDMEAVSCISIGVLDIFGFEDFERNSPEFCINAVNEQLOYFTQHFILKEQEYGO 562  
QY 475 EGIWMDIMDINGECGLDLEK-LGLLALINEESHFPQATDSTLEKLSQAHNHFY 533  
DB 563 EGISWHNIDYTDVNGCIHLISKPKTGLFYLLDESNFPHATSHLLAKFKQOHEDNKYFL 622  
QY 534 KPRVAVNFGVHAYGEVQYDVAGILEKNDFRDLNLNLRSESRDFIYDLF----- 586  
DB 623 GTFVLEPAFIHQFAPRVKQIKDFREKMDYMRPDIALLKSGSDSYRQGLGMPVAV 682  
QY 587 -----EHVS----- 590  
DB 683 FRWAVLRAAIRAMAVLREAGRLRAERAKEAGVSSPVTRSHVEELPRGANTPSEKLYRD 742  
QY 591 ----- 590  
DB 743 LHNQIIKSLKGLWGEDPRRLLOSLRLQKPTFLKSGIKQIKOQIIPKNLDSKSLR 802  
QY 591 -----SRNNODTLCKGSKHRRPTVSOPKVDLSLMAATLSSNPFVRCIKPMQMK 643  
DB 803 IISMTLHRTTKSLHLHKKKPPSISAQPT-SLKKLLEALCAKAPFTRICRSNAEKK 861  
QY 644 PQDFDAVVLNQLRYSGMLETYIRKAGYAVRRPQDFYKRYKVLNRLNALPEDV--RG 700  
DB 862 ELCFDDELVLQGRYTGMELETYIRRSYSAKYTFQDFTEQOVL-----LPDVOPCHE 916  
QY 701 KCTSLIQLIDANSSEMOGLTKYFLRESLEQLEKREBEV----- 741  
DB 917 AIALLEKLOVDRQNOYIGTKVFLKETERQALQERLHGEVLARRILLLOSWFRMVLRRH 976

QY 742 ----SHAAMVIRAHVGLFARKQYRKVLVCVILIQKNYAFLLRRRLHLKKAIVFOKO 797  
DB 977 FVQMKHAALTICACRSYVRRTLERT--RAAVYLLAARVGLQGRAYHHQRISIRLOSL 1035  
QY 798 LRQIARVRYROLAEKQOEKKOEBEKKRE-----EEREREREREA---- 845  
DB 1036 CRGHLORRSFQOMMEKQAEQARETAGAEMSEGPSPVPAAGEQSEHVEDEPESIGVET 1095  
QY 846 -----ELBAQOEERTKQOELEALQKSOKAEALFRELEKOKENKOVEI 889  
DB 1096 EFMWNSKSPNGISPKKEI-PSPEMETPAOKTVPA--ESHKVPSSREKRESRRQGLEIV 1152  
QY 890 LRLEKEIEDLOMKEQOELSLTEASLOKLOERDDELRLREBACRAAQEFLSLNFDPI 949  
DB 1153 EKQNHIGSCR-----BENSLRRPSKASLETGES--FPE- 1186  
QY 950 DECVNRIERSLSGSEFSEELAESACEEKNFNFSQPIPEEVEDGEFEADDDAFQDSPNP 1009  
DB 1187 -----DTKEPRREDGLETWETETAPSCQVPIVGDPP-----RSPSP 1223  
QY 1010 SEHGHS-----DORTSGIRTS-----DSSSEEDPYNNDTV--PTSPSADS-----TVL 1051  
DB 1224 LQRPASLIDSKSVFVLPSSLSLESPQDEBKGNSTKVODKPSPEGSTQIQRYQHPDTER 1283  
QY 1052 LAPSYOD-----SGSLHNSSGSESTYCMFQNAQDLPSPGPDYDQDDYEDGAI 1100  
DB 1284 LATAVIEMRGKTLASAMLSQSLDSEK-----PRTAGAALT----- 1320  
QY 1101 TSGSVTPNSYSGSQSPDYRCVSTYNSGAYRPSSEGAQSSFPDSEDFSRDPTDE 1160  
DB 1321 TEERISIFSTSDVSKLSP--VKISTEVDGDLAKKPPGHNKKKSPDPSGAPGACPTGSO 1377  
QY 1161 LSYRRDSVSCVTLFPHSFLMKGMLMSWRKRCVYLKDEFLMFRSQOZALQKQWLHK 1220  
DB 1378 -----GDSKSAFGR-----LFLHAKDKKPSLGEVET 1405  
QY 1221 KCGGSSTLSRRNWKRMFVLRQSKLMPFNDSSEKILKG-----TYEVRTAKEII 1269  
DB 1406 EBSGQQAQEAARATLDV--PSSQOHNHTTGEKLLKQKNNRNVGQITVSEKKRESVF 1463  
QY 1270 DNTYKENGIDITMADRTFHLIAESPEDASQMFVLSQVHAUSDQEIQEMHDEQANPONAV 1329  
DB 1464 RKITYANAEKFL-----DEFILNKVNDLRSQKPIESLFEATERFRSNYKTYSVN-- 1516  
QY 1330 GTLDVGLDSVCASDPRPNSFVITITANRVL-HCNADPREMHMITLLO--RSKQPT 1385  
DB 1517 GKIHVGKDLM-----ENYQIVSIVLAERGEKQDNLVNLVFOSLDEFTSYNKT 1567  
QY 1386 RVEGQEFIVRGWLHKEVKNSPKMSLSLKRWFLVLTNLSLDYKSEKVALK----- 1437  
DB 1568 DEEP-----VKG-----KAQKKRQKQBAVQEHNG-HVFASYOVNVPQSCQCLSY 1612  
QY 1438 ---LGTVLVNSLCSVPPDEKIFKETGYMNTVYRKHCYRLYTLNLAEATMSSVIONV 1494  
DB 1613 IWLMDKALCSYCKMTCHKKCVHKIQSYCSYT--GRRK-----SELGAEPGHFVCVSDSL 1665  
QY 1495 TDTKAPIDPTPOQLQDIKENCLNSDVQIYKRPILRYTHRPLHSPLLPYGDIMLN 1554  
DB 1666 TSDKASVPIVEKLELHEVMBGL--YTBGLFRKSGAANRFR-----ELROA 1709  
QY 1555 ILKDKGYTTLQDEALIKIFNSL--QQLSEMSDPIPIIOGILQTHDLR-----PLRDE---- 1604  
DB 1710 LOTDRTATYKLEDFPHAITGVAKQVRLRELPEP--LMPFAQVGDPLRAVELBEKEQDLAA 1766  
QY 1605 LYC-----OLIKQTNKYPHPOSVGNLYSWOILTC--LSCTFL 1639  
DB 1767 IYAVLDHLPEANHTSLERLIFHLVYKALLLEDVNRH--SEGALAIIPAPCILLRPRDSDPLT 1825  
QY 1640 PSRGILKYLKFLKRIEQPFCTEMEKALFTYESLXKTKCEFPVPSR 1687  
DB 1826 SMKDVAKITTCVEMLIKO-----MRKYV-RMEINHLAEASIAFR 1867



FT CONFLICT 1484 1484 E -> D (IN REF. 5).  
 SQ SEQUENCE 1855 AA; 215419 MM; 1C55AD57285FA9EC CRC64;  
 Query Match 13.7%; Score 1466; DB 1; Length 1855;  
 Best Local Similarity 25.7%; Pred. No. 9.1e-60;  
 Matches 517; Conservative 349; Mismatches 681; Indels 468; Gaps 78;

QY 7 RWLRENGOHFPSTVNSCAB-----GIWFRPDYGVFTYKQSTITHQVTAMHPT 57  
 DB 12 RWIIPD-----PEWVMSAELLDKDYKPGDKVLLHLHEEGKDLHLDPKTELPHLRNP 66  
 QY 58 NEGVDMASTLHGSSIMYNYFQYKRNQ-ITWTIGSLASVNPQIAGLYEPATME 116  
 DB 67 ILVGEVDLTLSTYLBPAVLNLRVREIDSKLIYTCGIVLVAINPEYQLP-ITYGEDJIN 125  
 QY 117 QYSRRHLGELPHIIPAIANECYRCIMWRHNOICLLKGESGAKTSTTKLLKFLSVISQ 176  
 DB 126 AYSQWQMGDNDPHIIPAVAEAYKOMARDENQSIIVSGESGAGKTVSAKAMRYFATVSG 185  
 QY 177 QSLLESLKEKTSCEVERALLESSEPIMEAFNAKTVNNSSRFGKFVOLNICOXNIQGR 236  
 DB 186 SASEANVEEK-----VLTSPIMESIGNAKTTBRDSSRFQKYTEIGFDKRYRIIGAN 238  
 QY 237 IYUCIISQNRVYRQNGERNYHIFALLAGLHEHEEEFYLTSPENYHILNOSGCVEDK 296  
 DB 239 MRYTL-EXRSVYFOAEERNYHIFYOLCASAKLPEFKMLRLGNADNFNTYTKQSGSPVIE 297  
 QY 297 TISDOSFREVIITAMDVMPFSKEEVEVSRLLAGILHIGNIEFTTAGQAVSF---KTL 353  
 DB 298 GVDADKEMATRAQCTILIGISESHQMGIFRILAGILHIGVGTSSRSDSCITPRKHBL 357  
 QY 354 GRSAGELGDPQLTALDQSRMFLRGEELTPLNVOQAVDSRSLMALVYACCFEMVIK 413  
 DB 358 CIFCDLNGVDYBEHCMLCHRLATATETETIKPISKLOATNARDALAKHAYALKFWMYD 417  
 QY 414 KINSRI-KGMBEDKSGIIDLIFGEPNFVNHFEOPNINANEKLOEYFNHIFSLQLEY 472  
 DB 418 NVAQALHASAVQKSHFGLVDIYGFETFEINSFEOFCINANEKLOOFPNHVFKLOEY 477  
 QY 473 SREGVWEDIDWIDNECDLIEKKGLALINEESHFPOATSDTLEKLSOHANN-HF 511  
 DB 478 MKQOITWTLIDFDNQCILWILSKIGIIDLDBECMFGKGTDTNAQKYNHLKCAL 537  
 QY 532 YKPRVAVNVFGVYHAGEVQDVRGILEKNRDTFRDDLNLRESRFPYIDLFEHVS 591  
 DB 538 FEKPRLSNKAFFIQHFADKVEYQCEGFLKKNKDTVEBOIKVLSKSFKMLPELF----- 592  
 QY 592 RNNODPLKCS-----KHRRPTYSSQPKVDSLHSLMA 623  
 DB 593 ---QDEKALISPTSATSSGRTPLTRPAKPTKGRPGQMAKHNKKTGVGHQFR-NSHLHME 648  
 QY 624 TLSSNPFVYRCIKPNQKMPDOFDOAVVLNOLRYSGMLETIVIRAGYAVRRPQDFYK 683  
 DB 649 TLMATTPHYRCIKPNDFKPFTEDEKRAVOQLRACGVLETTIRISANGPFSRTTYOEFSS 708  
 QY 684 RYKVLMENTALPEVDGKCTSLLOLYDASNEWOLGKTQVFLRESLEOKLEKREBEVSH 743  
 DB 709 RYKVLKQKQVLDSDRKQTCGNVLEKLIIDKQYQFGTKIFPRAQVAVYLEKLRADKLA 768  
 QY 744 AANVTIAHVGLFARQYKRVLYCV-----ITQKYRAP 778  
 DB 769 ACIRIKTIRGMLRKKYLMRAAITMORVYRGYARCYAKFLRTKATIIIOKWMRY 828  
 QY 779 LIRREPLHKKAIIVOKOLQIARVYQOLAE-----KXQES 819  
 DB 829 VARRKIRRAATIVQSYLRGFLARNRKIKLREHKAIVIQKVGWGLARTHYKSMHA 888  
 QY 820 -----KKQDEEKKKRKEEERER-----ERREALRAQOE----- 853  
 DB 889 IIVLOCFRMAKRELKUKIETARSVERYKLIHGEMKNIMQLOKRVDEQNDYKCLVE 948  
 QY 854 -----ETRK-QOELEALQKQKAEALT-RELEKQENKQVEEILRLKEIEDU 899

DB 949 KLTNLEGVNSFTEKLSDLERLQSEBEAKVATGVLISLQ-----EEIANKRLDLE-- 1000  
 QY 900 QRMKEQOELSTAEASLOKQERRDOELRLSEACRAAOEFLSJN---FDEIDECVRNI 956  
 DB 1001 ---QTRSEKCCIEBHADRKYQETQVLSUKENTLLKQ-KEALNHRIVQOAKETEM 1056  
 QY 957 ERSLSGSEFSSSELASACEKPNF---NFSQYPEEVEDEGFADDAFKDQSPNPE 1011  
 DB 1057 EKKL---VEETQLELDLNDERLRYQNLNFSR--LEERYD-LKEEWTLMHVHPK-- 1108  
 QY 1012 HGHSDORTSGITSDS-----SEEDPYNDIVPPSPASDVVLAPV----- 1056  
 DB 1109 -GH-KRTDSTHSSNESEYIFSSSELAEMEDISRTREEPEKKVPLDMSLFLQKRVTEL 1165  
 QY 1057 -ODSGSLHSSSGSESTYCWPNAGDLPSPD---GDYD---ODDYEDGATISGSSYVF 1108  
 DB 1166 EDEQVMDLDRKEQVLRSKAKKEERQIIGALEFYSLQKQLESNKKLKNEML 1225  
 QY 1109 SNSYGSQWSPDYRCVGYNNSGAYRFSSEGAQSSFEDEEDFDRFDTDELRYRDSV 1168  
 DB 1226 RALSEKSAPEV---TAPGAPAYRVLMEQLTS-----VSEELDVRKEEV 1266  
 QY 1169 YSCVTLPHYHSLFYMKGLMNSW-----KRRVCYLDKDFILM-----FRSQBALQOGL 1218  
 DB 1267 -----LILRSQVSKQKALIQPDKNMTDSTILLEVDQMKDKQ-IAQAYI 1313  
 QY 1219 H-KKGGSSSTLRRMWK---KRMFYLR-----QSKLMYFENDSEKLGTYE 1261  
 DB 1314 GUKETRSSALDYHELNEGELMVLVEGKQANRLLESQLOQKXSHENEA-EALRG-E 1370  
 QY 1262 VRTAKEIIDNTTENGIDITIMADRTPLIABS---PEDASQWPSVLSQVHASTDOETQEM 1318  
 DB 1371 IOSLKE-----EN-----NBOOQLAQLQVPEARIBASIQHEITRLTNEMLDLM 1416  
 QY 1319 HDEQANPN-----AVGTLVDLISVCS-----DSPDRNSVIIITANV 1360  
 DB 1417 --EQLEKQKTVRKLLQKQVPAKKIGELEVOQMENISPGQIIDEPIR----- 1463  
 QY 1361 LHONADTPEEMHMTLLQORSKQDTRVEGEQFIVRGMLHKEVNSPKMSLKLKR----- 1416  
 DB 1464 ---VNIPKEDPQGMLEYKEDQ-----KLVN-----LILELKRGVAV 1502  
 QY 1417 -----WFLTHNSLDYKSSKQNALKGLVLVNSLCSVVPDEKIFKETGYMV 1465  
 DB 1503 NLIPLGPAVILPWCVRH--ADVLNDQK-VRSLLTSTINSIKKVLKRGDDPEVFSFWS 1559  
 QY 1466 TYVGRKHCYRLTK---LINAETRMSSVIOQVNTOTKAPIDPFTQOIODIKENCIN--- 1518  
 DB 1560 NTCRFLHCLKQYSGEGBFKNHTSR---QN-----EHCULTNPD 1594  
 QY 1519 -----SDVVEQIYKRNPIRLRYTHHPLASPLP--LPYGD1-NILNLLKDKGY---TT 1563  
 DB 1595 LAEYQVSLDALQIYQO--LVRLVENILOPVIYSGMLEHETIQGVSGVKPGLKRRSS 1652  
 QY 1564 LODDAIKIFNS-LQOLESMSDPIPIIOGILQTHDLRPLRDELYQOLIKQTNKVPHPGSV 1622  
 DB 1653 IADEGTYTLDSTIRLONSFHS-----VMCOHQMDP--ELIKQVWQOMYIIGAITL 1701  
 QY 1623 GMLYMQOILTCSCTFLPBRGILKYLKFLHKIRIQFPQTE-MEKKALFTYESL----- 1675  
 DB 1702 NNL---LIRKDMCSW--SKG--QIRYVNSOLEMRLDNKIMNSGAETLEPPLIOAOL 1753  
 QY 1676 ---KTKCREFPVPRDEIEALIHROEMTSYVYCHGSGSCKI--TINSHTTAGEVEKUL- 1728  
 DB 1754 LOVKKTD-----DDEALCS-----MCAALTTAQIVKVLNLATTPVNEBERVS 1797  
 QY 1729 ---IRGLAME-----DSRNMFL-FEYN 1747  
 DB 1798 VSFIRTIQWRLDRDPSQPLMDAKHIFVTPFPFN 1832

RESULT 13.  
 MYSC\_HUMAN

ID	MYSC_HUMAN	STANDARD;	PRT; 1742 AA.
AC	Q9NOX4;		
AD	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Myosin Vc (Myosin 5C).		
CN	MYOS5C.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.		
RX	MEDLINE=21859382; PubMed=11870218;		
RT	Rodriguez O.C., Cheney R.E.;		
RT	"Human myosin-Vc is a novel class V myosin expressed in epithelial cells";		
RU	J. Cell Sci. 115:991-1004(2002).		
CC	- FUNCTION: May be involved in transferrin trafficking. Likely to		
CC	power actin-based membrane trafficking in many physiologically		
CC	crucial tissues.		
CC	- TISSUE SPECIFICITY: Expressed chiefly in non-neuronal tissues.		
CC	Particularly abundant in epithelial and glancular tissues		
CC	including pancreas, prostate, mammary, stomach, colon and lung.		
CC	- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.		
CC	- SIMILARITY: CONTAINS 6 IQ DOMAINS.		
CC	- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; AF272390; AAF78783.1; --		
DR	HSSP; P08799; 1VOM.		
DR	Genew; HGNC:7604; MYOS5C.		
DR	InterPro; IPR002710; DIL.		
DR	InterPro; IPR000048; IQ_region.		
DR	InterPro; IPR001609; myosin_head.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	Pfam; PF00612; IQ; 6.		
DR	Pfam; PF01843; DIL; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	Prodom; PD000355; myosin_head; 1.		
DR	Prodom; PD003576; DIL; 1.		
DR	SMART; SM00015; IQ; 5.		
DR	SMART; SM00242; MYSC; 1.		
DR	PROSITE; PS00096; IQ; 4.		
KW	Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;		
KW	Coiled coil; Polymorphism.		
FT	DOMAIN 1 755		MYOSIN HEAD-LIKE.
FT	DOMAIN 756 779		IQ 1.
FT	DOMAIN 780 806		IQ 2.
FT	DOMAIN 807 829		IQ 3.
FT	DOMAIN 830 854		IQ 4.
FT	DOMAIN 855 884		IQ 5.
FT	DOMAIN 884 1351		COILED COIL (POTENTIAL).
FT	DOMAIN 1574 1679		DILUTE.
FT	VARIANT 522 522		P -> L.
FT	VARIANT 522 522		L -> S.
FT	VARIANT 634 634		/FTID=VAR_010646.
FT	SEQUENCE 1742 AA; 202793 MW; 25DD3082A7B5AAB CRC64;		/FTID=VAR_010647.
QY	Query Match 13.5%; Score 1448.5; DB 1; Length 1742;		
	Best Local Similarity 27.0%; Pred. No. 5,3e-59;		
	Matches 457; Conservative 298; Mismatches 613; Indels 327; Gaps 49;		
	61 GVDMA5LTELHGSIWNLFORVYRNO-IMTYIGSILASVNPYQPIAGLYEPATMEQYS 119		

Db	68	GENDTLSTYLHEAVLHNIRAREBKLITYTSGIILLVAMNPYKOLP-IYGDALIHAYS	126
OY	120	RRLHGEELPHIFALANECYRCLMRHNOCLIKGSGAGKTBESTKILNLFVLSYQOOL	179
Db	127	GOMNGMDPHIFAAVEAAVYKOMARNNNQSIIVSGSGAGKTAVAFAMRFATVSKGS	186
OY	180	ELSLKERTSCYERAILLESSPIMEFNAKYVYNNSSRFCKFVOLNIQCKNIOGGRIVD	239
Db	187	NAVEDK-----VLASNPIITVEVNAKATTRDNSSRFQYTEI8FDEONQIIGAMNST	239
OY	240	CISSORVVRONPGENRYHIFVYLLAGLEHEEREFPYLSPTENYHILNOSGVEDTIS	299
Db	240	YLL-EKSRVVFQSENERNYHIFVOLCSAQOSEFKHLKLSAEFPNTTRMGNTVIEGVN	298
OY	300	DOESFREVITAMVOPFSKEEVRVSHLAGLHAGNIBFTAGACVSPK--TALGRS	356
Db	299	DRAMVEITQKFTLLGFPEBDQOMOVFIILAIHLGNVOQTAVGNERSVSEBDSHLKVF	358
OY	357	AEILGLDPTQLTDLTORSPFLRGEELITPLANOAVDSRDSLAMALYACCFEYVIKIN	416
Db	359	CELLGESGRVAQWLCNRKIVTSETEVWKPMTRPOAVNADALAKKIYAHFDPIVERIN	418
OY	417	SRIK-GNEDPFSIGILIDFEPENEVNHPEOFNIVNANEKLOEYFNHHSFDELEKRE	475
Db	419	QALOPSGQHTFIVLBIYGFETDVNSFEOPCINAVNEKLOQOFNHFVLEBOEBEYKE	478
OY	476	GLVEMIDIMDNGECLDILIEKUGLILLINESHFPDATSTLTLEKLSHQAN-NHFYVK	534
Db	479	DIPTTLIDFYONGVIDILIEAKMGIIBLBDECLPHGTDENMPOKYVNNPVNBNPLPEK	538
OY	535	PRVAVNNGVKHVAGEVOYDVRGILEKNORDTFRDDLNLRESFDFIYDLFEH-----	588
Db	539	PRMNTSGFVIOHFPADKYEYKEGFELEKRDVTYMLVEIIRASGFPHCANFQENPPTPS	598
OY	589	-----VSSRNNQDILKGGSKHRRATVSQFNVDSHSLMTLSSNPFWRCIDENNOXM	643
Db	599	PGSGMITVYKSAQVIKPNRSKHFRTTVASKPE-SSLYILMETLANATTHYVNCIPNDEKL	657
OY	644	PDOEPDVAVLNLQLYSGMLFETVIRKAGYARPRFODFYKRYKYLAMENLALP-EDVRGKC	702
Db	658	PFEEDSKRIVOQLAACVLETRISASQYPERWYIEFYSRYGILMTKQELSPSDKKEVC	717
OY	703	TSLLQLYDASNEWOLGKTQVFLRESLEQLEKREBEVSHAAVIRAHVGLFARQY-761	
Db	718	KVVLHRLIJDSONOYQFGTKIIFPAGOVAYLEKRLDKLRQS CVWQKHMGMLORRKFL	777
OY	762	-----RKL-----YCVIITQKYRAFLLRRFLHLKKAALVF	794
Db	778	RERRAALIOQYFRGOOTVRKAITAVALKEAMAIIOKRCRGVLSYOLIMATITM	837
OY	795	QKOLRGQIARVRQOL-----LAEKREOEKK-----KOEBEKK	830
Db	838	QAYSBGFLARRRYKMLEBKHAVLLOKYTARAMLRRRPOSTRRVVLNIQLTFRVORLOKK	897
OY	831	REEERER-----EERREAEILRAOEESTRKOQELER-LQSQEAEILTRLEKOKENKQ	885
Db	898	LEDQNKENHGVLEKLTSLAALRAQDVE---KIQGLAELEKAAVHRRNYEKGGRYDAY	954
OY	886	VEILIRLEKEJEDLQRMKEQOELSLETSLOKLOERPDQIBRLREBEACRAAOEFLSLN	945
Db	955	EELKALQKHNSELETOKEQIOLTKOERT-BELKEKMDNLTKOL-----997	
OY	946	FDEIDECFRNIERSLSGGSFSSSLASACEKNFNFQOPYPEEYVDEGEADDDAFKD	1005
Db	998	FDDOVK-----EERQMLLEKSFELK-----TOY--EKQIQSLKEBITKALKD	1038
OY	1006	SPNSESQHSDORTSGIRTSDDSEEDPYMNDVTVPSPASDITVLLAPVSODGSLHNS	1065
Db	1039	EKMOLQHLVEBENH-----TSDGLAEAVARLSKVKITSEFEKEIHELLQAKQIDVBKHYQS	1094
OY	1066	SSGESYTCMPONAGDLPSPDGDYDYDDDYEDGAIITSGSSYTFNSNSYGQWSPDYRCSGV	1125
Db	1095	QKREBREMSEITKOLLE-----SYLIED-----VRRLSYE-----DLE	1129





```

Db 358 IIFCDJMGVDEEMCHMLCHRKLATATETTYIKPISKLOATNARDALAKHIVAKLFNWIWG 417
Qy 414 KINSRI-KGNEDEFKSIGLIDIFGFENFENVHFEQFINANERKLOEFNKHIFSEOLEY 472
Db 418 HYNOMLSHVAKOSHPIGLVDIYGFETFEINSEFOFINANERKLOEQFMHVFKEQBEY 477
Qy 473 SREGLWEDIDWIDNGECLDLIEKLGALLAINESHESFOATDSTLLEKXHSQHANN-HF 531
Db 478 MKEQIPWTLIDFYDNQPCINLIESKLGILDLDEBCKMEFGTDPTWAQCLYTHLANKAL 537
Qy 532 YKPRVAVNFEVKNYAGVOYDVNGILAKNDTPRDDLLNLRESRPFIYDLFEH--- 588
Db 538 FEKPRMSNKAFLIKHAPADVEYOCSEGFLEKNKDYEEOJ KYLKSCKFKMLPEQDDCK 597
Qy 589 -----VSSRNQDTLKCSGKRRFPYVSQFKDLSHSLMATS 626
Db 598 AISPTBATSSTGPTLTVPVKPTKGRPGT-----AKEHKTVGLQFR-NSLHLMETIN 651
Qy 627 SSNPFVRCIKENMKMPDQFOAVVLANQARYSGMLTIVIRKAGYAVRRPQDEYKRYK 686
Db 652 ATPHYVRCIKENDFPFTFDBEKRAVQQLRAQGVLETIRHISAGFPSMTYQEFPSRYR 711
Qy 687 VLMKRLALPEDVGRGKCTSLDLYDASNSMOGKTYVFLPESLEQLEKREBEVSHAM 746
Db 712 VLMKQDVLGDRKQTCQVLEKILDKQYQFKTIFPAGQVAYLEKLRADKLRACI 771
Qy 747 VRAHVLGFLARKO-----YRKVL---YCVVLIQKNVAFILR 781
Db 772 RIQKTIKGLKRLKRYLCMQRAAITVQRYRYGYOARCYAKPLRRTKAATTQKXMYRYVR 831
Qy 782 RRFLLHKKAAIVFOKOLRGQIARVYRQLAE-----KREOE--- 819
Db 832 RYKTKRRAATIVLYGSLRGYLAHRKYLREHKAVTIQKRYVGMARHTHYKTKAILY 891
Qy 820 ----KKQEEBEKKKKREERER-----ERRAHLPAAQOE----- 853
Db 892 LQCCFRMMWAKRELKLIKARSVERKYKLIHGMENKIMQLQKQVDEQKDYKCLMEKLT 951
Qy 854 -----ETRK-OOELALOKSOKAEELT--RELEKOKNNOVEILREKEIEDLOM 902
Db 952 NIEGYVNSTEKLKNDVERLOJSEBAKATGKIVLSQ-----EELAKRDLDETRSE 1005
Qy 903 KEQOELSLTEASLOKQERRQDELRLREE--ACRAQOELSLNDEIDECEVNIERSL 960
Db 1006 KKS-----IEERADKYQKETEQLVSNLKXENTLLKQEKMLHLMWEQKEMTERMERTL 1060
Qy 961 SGGSEFSESLAESACEKPNF-----NSQPYPEEVEDGEFEADDAFKOSPNSEHGHS 1015
Db 1061 ---VETIKOLELDLNDERLRYQNLNIEFSR--LEERYDD--LKEEMTLMINVKRP--GH- 1110
Qy 1016 DORTSGIRTSDDS-----SREDPYMNDTVPTSPASDSTVJLAPSVQDSGLNSSGSEST 1071
Db 1111 -KRTDSTHSSNSEYTFSSSEPAETEDIARTBEPTEKKYPLDMSLFLKLOKRVTELQOK 1169
Qy 1072 YCMPQNAQDLPSPDGDYDDODDYEDGATTSGSVTFSSNSYSGOWSPDYRCVGTYSNGS 1131
Db 1170 QLM-----QDELD---RKBEQVLRKSKKGE--RPOKRGHMGJESJLK 1207
Qy 1132 AYRFSEGAQSSFDESE-----EDPSPRDTDELSTYRD 1166
Db 1208 ROELESSENKLLKNELEMLKALSEKAPRVNAPGAPAYRVLMEQLTA---VSEELDVREX 1264
Qy 1167 SYVSCCTLPYFHSPLVMKGLNMSW-----KRWCLXDEETLPMFSSKQKALKQGLHKK 1221
Db 1265 EV-----LILRSQVLSQKEAIPKODKNTMTDSTIL-LEDVKKMKDKGFIADA 1311
Qy 1222 GGGSSITLSRRNWKRFVLRQSLMYFENDSE---EKLKATVEVTRAKIILNNTKENGII 1278
Db 1312 YIGLMEKN-----LLESQLOQSKRSHENEMALNR3--EIGSLKE-----EN-- 1351
Qy 1279 DIIIMDRTHLIAES---PEDASQMFVSUQVHASTDOELQEMHDEQANPON----- 1327
Db 1352 -----NRQOOLLAONQLPPEARIEASLOHEITRLTNEMLDIM--EQLEKQKTVYKLLK 1404

```

```

Qy 1328 -----AVGLDVLGIDSVCS---DSPPBNSFVITANKVILCHNADTPEEMHWITL 1377
Db 1405 QUKVFAKKIGELFVQGMENISBQIIDEPIR-----VNIRPKGKDPOGM 1449
Qy 1378 LORSKGDTRVEGQEFIVRGMLHKEVNSPKSSSLKTKRWFLVLTNS----- 1424
Db 1450 LEYKREDO-----KLVKN-----LILELKRGVAVNLISGLPAYILPMCVRH 1492
Qy 1425 LDYKSSSKNMLKGLTVLNSLCSVPPEDEKIFKEXTGYMNVTYVGRKKCYRLTK-----L 1480
Db 1493 ADYLDQDK-VRSLLTSTINSIKVLRKRGDDFETVSWLSNTCRFLHCLQKQSGEBGM 1551
Qy 1481 LNEATRWSSVIONVTDTRAPIDTPQQLIODIKENCLN-----SDVEQIYKKN 1529
Db 1552 KHNTR-----QN-----EHCINPDLAEYRQVLSDLAIGIYQO- 1585
Qy 1530 PILRYTHHPLHSPILP--LPYGD-I-NLNLKDKGY---TLQDEAIKIFNS-LOQLBSM 1581
Db 1586 -LVRLENILQPMIVSGMLHEHTIQVGGVKPTGLRKRTSSIADGTYTLDSILKQLNSF 1644
Qy 1582 SDPIPIIGILOTGHDRLAPLDELXCOLIKOTNKYPHGSVGNLYSQULITCLSTPLBS 1641
Db 1645 HS-----VWCQGMDB--ELIKQVAKQMPYIVGAILTNL--LLRDMQSW--S 1687
Qy 1642 RQILKYLKFLKRIEOPFTE-MEKVALFTYESL-----KTKCRFPVPSRDEI 1690
Db 1688 KCM--QIRVNSQLEBMLRDKLMSGAKTELEPILOAQLOVQKTD-----DDA 1737
Qy 1691 EALIRQEWSTVYCHGGGCKI--TINSHTAGEVERL-----IRGLAME----- 1735
Db 1738 EATCS-----MCNALTTAQIVKVLNLYPVNEFEERVSVPFRTIQVRLDRKDSPO 1789
Qy 1736 ---DSRNMFL-FEYN 1747
Db 1790 LMDAKHIFPVTFPPN 1805

RESULT 15
MYS2_YEAST
ID MYS2_YEAST STANDARD; PRT; 1574 AA.
AC P19524;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin-2 isoform.
GN MYO2 OR CDC66 OR YOR326W OR O6167.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF98;
RA MEDLINE=91201404; PubMed=2016335;
RT Johnston G.C., Prendergast J.A., Singer R.A.;
RT "The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin
RT for vectorial transport of vesicles."
RL J. Cell Biol. 113:539-551 (1991).
RN [2]
RP SEQUENCE OF 1-748 FROM N.A.
RC STRAIN=628BC / FY1679;
RA MEDLINE=97051589; PubMed=8896266;
RT Pearson B.M., Hernandez Y., Payne J., Wolf S.S., Kalogeropoulos A.,
RT Schweitzer M.;
RT "Sequencing of a 35.71 kb DNA segment on the right arm of yeast
RT chromosome XV reveals regions of similarity to chromosomes I and
RT XII."
RL Yeast 12:1021-1031 (1996).
RN [3]
RP SEQUENCE OF 677-1574 FROM N.A.
RA MEDLINE=97051586; PubMed=8896263;
RT Parle-Mcdermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;

```

RT "Sequence of 29 kb around the PDR10 locus on the right arm of  
RT Saccharomyces cerevisiae chromosome XV; similarity to part of  
RT chromosome I."  
RL Yeast 12:999-1004(1996).  
CC -1- FUNCTION: ESSENTIAL MYOSIN FOR VECTORIAL TRANSPORT OF VESICLES.  
CC PLAYS A DETERMINANT ROLE IN THE SPATIAL REGULATION OF BUD  
CC FORMATION.  
CC -1- SUBUNIT: MAY SELF-ASSOCIATE. INTERACTS WITH MLC1.  
CC -1- MISCELLANEOUS: CA(2+) MAY PLAY A ROLE IN MYO2 ACTIVITY.  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M35532; AAA34810.1; -  
DR EMBL: X30565; CAA62184.1; -  
DR EMBL: 249821; CAA89973.1; -  
DR EMBL: 275234; CAA9646.1; -  
DR EMBL: 275235; CAA9646.1; -  
DR PIR: A38454; A38454. -  
DR HSSP: P10587; 1BR2. -  
DR SGD: S0005853; MYO2. -  
DR InterPro: IPR002710; DIL. -  
DR InterPro: IPR000048; IQ\_region. -  
DR InterPro: IPR004009; Myosin N. -  
DR InterPro: IPR001609; myosin\_head. -  
DR Pfam: PF00063; myosin\_head; 1. -  
DR Pfam: PF00612; IQ; 6. -  
DR Pfam: PF01843; DIL; 1. -  
DR Pfam: PF02736; Myosin N; 1. -  
DR PRINTS: PR00193; MYOSINHEAVY. -  
DR ProDom: PD000355; myosin\_head; 1. -  
DR ProDom: PD003376; DIL; 1. -  
DR SMART: SM00015; IQ; 2. -  
DR SMART: SM00242; MYSC; 1. -  
DR PROSITE: PS50096; IQ; 1. -  
DR Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation;  
DR Multigene family; Phosphorylation; Calmodulin-binding.  
KW Multigene family; Phosphorylation; Calmodulin-binding.  
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.  
FT 832 861 IQ.  
FT 926 981 COILED COIL.  
FT DOMAIN 982 1009 HINGE.  
FT DOMAIN 1010 1086 COILED COIL.  
FT DOMAIN 1087 1574 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.  
FT 443 523 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 750 925 ARG/LYS-RICH (BASIC).  
FT 791 805 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 814 828 CALMODULIN-BINDING (BY SIMILARITY).  
FT 839 853 CALMODULIN-BINDING (BY SIMILARITY).  
FT 862 876 CALMODULIN-BINDING (BY SIMILARITY).  
FT 887 902 CALMODULIN-BINDING (BY SIMILARITY).  
FT 910 924 CALMODULIN-BINDING (BY SIMILARITY).  
FT 164 171 ATP (BY SIMILARITY).  
FT 692 692 ALKYLATION (BY SIMILARITY).  
FT MOD RES 702 702 ALKYLATION (BY SIMILARITY).  
FT MOD RES 1097 1097 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
FT 1574 AA; 180680 MW; 1F7E2887C1E59D54 CRC64;  
SO SEQUENCE

Query Match 13.4%; Score 1442; DB 1; Length 1574;  
Best Local Similarity 32.5%; Pred. No. 9.2e-59;  
Matches 311; Conservative 228; Mismatches 413; Indels 130; Gaps 29;

QY 55 HPNKGVDMDMLTELHGSSIMYNLFQYKRNQITTYIGSILASVNPQPIAGLEPPT 114  
DB 65 NPEILATEDLSTLSYLNBAVLAHKORYSOLNITYTSGIVLIATNPDRDQLYTQDM 124

QY 115 MEQYSRRLGELPRHIFALINAEVCYRCLWRHNDQILKGSAGKTESTKLILK-FLSV 173  
DB 125 IQAYVGRGELPEHLFAIAEAYRLMKNQDNQITVSGSGACKTYSAKITMYPSV 184  
QY 174 IQQSLSTLSEKTECEVERAIISSPIMEAFGNATVYNNSSRGKVFQNLICQKINQ 233  
DB 185 EENENATVOHQVEMSETEQKILATNPIMEAFGNATTRDNDNSRGKYLEILFDQDTSI 244  
QY 234 GGRIVDCILSSQNRVVRQNGERNYHIFALLAGLEHERREFFVSTPENHYALNQSGCV 293  
DB 245 GARITTYLL-ERSRLVQPPRIERNYHIFQLMAGIPACQKEHLHLDASDYVYMMQSGT 303  
QY 294 ERTISDQSFREVTANDVWQFSKEVEVSRLLAGILHGNIEF-ITAGAQVSF-KT 351  
DB 304 KINGIDDAKEYKITADLTALVIGITKETOHQIKLIALHIGNIEIKKTRNDASISADBP 363  
QY 352 ALGRSAELLGLDPTQTLALTRSMFLRGEELITPLANQOAVSDSLMALIACCFEV 411  
DB 364 NLKLCELLGIDAYNPAKVTYKQITRSEKIVSNLNSQALVAVDSVAKFTYSALFDWL 423  
QY 412 IKKINSRIKG---NDEPKS-IGILDIQGFENFEVNHFEQFINVYANEKLOEYFNKHIFSL 467  
DB 424 VENINTVLCNPAVNDQISSFIVGLDIYGFHEHEKKSPEQFCINVANEKLOEFPNGHVFPL 483  
QY 468 EQLSEYRGLVWEDIDWIDNGECLDLEKQGLLALINEESHFPQATDSTLLEKLH---S 524  
DB 484 EQLSEYRGLVWEDIDWIDNGECLDLEKQGLLALINEESHFPQATDSTLLEKLH---S 524  
QY 525 QHANNHFYKPRVAVANNQGVYKAYAGEVDYRNGILEKXNDTRDDLNLBRSRDFIYD 584  
DB 544 KSPRTKVSFKPRFGQTKFVSHYALDVAVDGFLEKNDIVSDHLEVLKASTETTLN 603  
QY 585 LFEHV--SSRNNQDTLK---CGSK-----HRRPTVSSQFVDSLSHSMATLSSN 629  
DB 604 ILEGLEKAAKLEBAKLELEBQAGSKKPPRTVNRKPLTSGMFK-QSILEMNTINSN 662  
QY 630 PEFVACIKENKQMPDQDAVVLNQLRYSGLMLETVRIRKAGYAVRPPQDFYKRYKVL 689  
DB 663 VHYICIKRNADKEMQFQNLMLVLSQLRACGVLETIRISCAQFSPRWTFEEVLRYYLI 722  
QY 690 -----RYLALPEYRGKCTSLDLQYDASNEMQLGKTVFLRESLEKLEKRR 739  
DB 723 PHEQWDLIFPKKETEEDDISVKNMILDATYKDKSKYQGNKIFPKGMALYLELGRSN 782  
QY 740 BYSHAMVIRAVLFGRLAKOYRKVLYCVIIOKYRPAFLRRFLHLKK-AAIVFQO 797  
DB 783 KQNSIWMIOKKIRAKYTRKQYLOISQAIKYIQQNNIKGFIIRQVNDENKVCATLLQA 842  
QY 798 LRG-QIARRVYQLAEKREBEKKOEBEKKREBE-----EREREREREAELR 848  
DB 843 YGHGISIRANVFVLRTITNLQKKIRKELKQRLQKQHEVNAAVTIQSKVTFEPSPRFLR 902  
QY 849 AQOE-----EETKQOELALQKS-----QKAEELTREL-EKQEK 884  
DB 903 TKKDTVVVQSLIRRAAQKQKQLKADAKSVNHLKESYKLENKYIELQNLASKYKKEK 962  
QY 885 -----QVEILRLKEIE-----DQWRKE-----QOELSTEASHQ 916  
DB 963 EMTETIKELQVVEASAKIQETLENMKKENHLIDINQSKMELQKTIENNIQSTEQTLK 1022  
QY 917 KIQ-----ERRQOELARLEE-EACRAOEFLESLND-----EIDECVANNIR 958  
DB 1023 DAQLELEMDVQKHDELKESKKQLELEBQTKTLVVEYQTLNDLQNEVSLKEELARLQ 1082  
QY 959 SLSGSSEFSSELAESAACEKPN---FNSQPIPEEVEVDEGFADDAFADSNPSE-NGH 1014  
DB 1083 AMSLGTIVTTSVLPQPLPLDWMGGASNFRNMWLE---NSDSLSPNDLNLKSRSTPSSGNH 1139  
QY 1015 SD-----QRTSGIRTDSDSEEDPYVNDPVPTSPSADSVLLAPSVQDSGLHNSGGS 1070  
DB 1140 IDSLSVDRENGVNAQINBELYRLLEDTEILNQELTEEC-LKGFVEVPDGAVALQUSKRDV 1198  
QY 1071 TY 1072

Db 1199 VY 1200

Search completed: July 14, 2003, 18:16:57  
Job time : 43.3952 msec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:07:56 ; Search time 95.4296 Seconds  
(without alignment)  
4441.379 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 PCLQCTRWLRNGQHFPST.....MIVKKRYSTTRSSAQSSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10011	93.3	2062	11	Q9JUY5 mus musculus
2	4744	44.2	924	11	Q8R360 Q8R360 mus musculus
3	3294	30.7	712	4	Q94893 Q94893 homo sapien
4	1921	17.9	2357	5	Q9UIM8 Q9UIM8 dicystosell
5	1843.5	17.2	2167	5	Q9VJZ6 Q9VJZ6 drosophila
6	1834.5	17.1	2179	13	Q9VJZ6 Q9VJZ6 drosophila
7	1831.5	17.1	2179	13	Q9VJZ6 Q9VJZ6 drosophila
8	1716	16.0	2098	5	P91443 P91443 caenorhabd
9	1705	15.9	2129	5	Q9VJZ3 Q9VJZ3 drosophila
10	1697	15.8	2121	5	Q9VJZ3 Q9VJZ3 drosophila
11	1635	15.2	2113	11	Q9VJZ3 Q9VJZ3 drosophila
12	1512.5	14.1	1529	10	Q9VJZ3 Q9VJZ3 drosophila
13	1495	13.9	2424	5	Q9VJZ3 Q9VJZ3 drosophila
14	1492	13.9	2548	4	Q9VJZ3 Q9VJZ3 drosophila
15	1487.5	13.9	1611	10	Q9VJZ3 Q9VJZ3 drosophila
16	1487	13.9	2626	11	Q9VJZ3 Q9VJZ3 drosophila

17	1485	13.8	1260	10	Q24516 Q24516 helianthus
18	1483	13.8	1528	10	Q24518 Q24518 helianthus
19	1480.5	13.8	1533	10	Q8S0C8 Q8S0C8 cryza savi
20	1466	13.7	1511	10	Q9M5A6 Q9M5A6 valianeria
21	1463	13.6	1547	10	Q8W312 Q8W312 cryza savi
22	1458.5	13.6	1477	10	Q64491 Q64491 arabidopsis
23	1458	13.6	1520	10	Q39160 Q39160 arabidopsis
24	1450.5	13.5	1556	10	Q9ZVW3 Q9ZVW3 arabidopsis
25	1449.5	13.5	1515	10	Q9ATB7 Q9ATB7 petroselinu
26	1441.5	13.4	2148	5	Q9BDJ3 Q9BDJ3 physarum po
27	1440	13.4	1505	10	Q9LKB9 Q9LKB9 arabidopsis
28	1437	13.4	1824	5	Q9NBH3 Q9NBH3 strongyloce
29	1435.5	13.4	1599	10	Q9LNP7 Q9LNP7 arabidopsis
30	1435	13.4	1515	10	Q39158 Q39158 arabidopsis
31	1434.5	13.4	1446	10	Q9SVT9 Q9SVT9 arabidopsis
32	1431	13.3	2651	10	Q9FRF5 Q9FRF5 arabidopsis
33	1430.5	13.3	1490	10	Q9SKB0 Q9SKB0 arabidopsis
34	1430	13.3	2167	10	Q9SSU1 Q9SSU1 chara coral
35	1430	13.3	2182	10	Q9LW97 Q9LW97 chara coral
36	1424.5	13.3	1242	10	Q9M2K0 Q9M2K0 arabidopsis
37	1423	13.3	1374	10	Q9SMY9 Q9SMY9 arabidopsis
38	1399.5	13.0	1502	10	Q24517 Q24517 helianthus
39	1399.5	13.0	1736	10	Q23025 Q23025 arabidopsis
40	1399	13.0	1516	3	Q94477 Q94477 echinococh
41	1394	13.0	1502	10	Q9SK73 Q9SK73 arabidopsis
42	1392	13.0	1285	4	Q9BZ27 Q9BZ27 homo sapien
43	1391	13.0	1285	4	Q9UEG2 Q9UEG2 homo sapien
44	1386.5	12.9	1304	13	Q9PWF6 Q9PWF6 morone saxa
45	1379	12.9	1471	3	Q74805 Q74805 echinococh

## ALIGNMENTS

RESULT 1  
Q9JUY5 PRELIMINARY; PRT; 2062 AA.  
AC Q9JUY5; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Myosin X.  
GN MYO10 OR MYO 10.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=20261293; PubMed=10799329;  
RA Yonezawa S., Kimura A., Koshida S., Masaki S., Ono T., Hanai A.,  
RA Sonta S., Kageyama T., Takahashi T., Moriyama A.,  
RT "Mouse myosin X: Molecular architecture and tissue expression as  
RT revealed by northern blot and in situ hybridization analyses.";  
RL Biochem. Biophys. Res. Commun. 271:526-533(2000).  
DR EMBL; AJ249706; CAB56466.2; -.  
DR HSSP; P08799; 1MND.  
DR MGD; MGI:107716; Myo10.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR001609; myosin\_head.  
DR InterPro; IPR000857; MYTH4.  
DR InterPro; IPR001649; PH.  
DR Pfam; PF00612; IQ\_3.  
DR Pfam; PF00603; myosin\_head; 1.  
DR Pfam; PF00784; MYTH4; 1.  
DR Pfam; PF00169; PH; 3.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR PRODom; PD000355; myosin\_head; 1.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00015; IQ; 3.  
DR SMART; SM00242; MYSC; 1.

```
DR SMART; SM00139; MYTH4; 1.  
DR SMART; SM00233; PH; 2.  
DR PROSITE; PS50057; BAND_41_3; 1.  
DR PROSITE; PSS0096; IQ; 1.  
DR PROSITE; PSS0003; PH DOMAIN; 2.  
SQ SEQUENCE 262 AA; 23746 MW;
```

Query Match	93.3%	Score 10011	DB 11	Length 2062
Best Local Similarly	93.3%	Pred. No. 0		
Matches 1925; Conservative	64	Mismatches	64	Indels 10; Gaps 6

Qy	1	CLQSTRWMLBNGOHFSPSTVNSCAEGLVPEPTDVGQVFTYQOSITTHOKRTAMHPTEE	60
Dd	4	FPFEGARWMLBNGOHFSPSTVNSCAEGLVPEPTDVGQVFTYQOSITTHOKRTAMHPLEE	63
Qy	61	GVDDMASLTELHGSGIMYNLFORYKRNQIWTYGSILASVNPQPIAGLYEBATMEOYSR	120
Dd	64	GVDDMASLTELHGVAIMYNLFORYKRNQIWTYGSILASVNPQPIAGLYEBATMEBYSR	123
Qy	121	RHGLBPHFIRAIANECTRCMLMKRDNCILIKSGSGAKTESTKULIKPLSVISQOSLE	180
Dd	124	CHGLBPHFIRAIANECTRCMLMKRDNCVILSGSGAKTESTKULIKPLSVISQOQTLD	183
Qy	181	LSLKEKTSYBRALIESSPIMEFAFNATKVYNNNSRSGKFLQULNICQGNIGGRIVDC	240
Dd	184	LGLOEKTSYVEQALIQSSPIHEAFGNATKVYNNNSRSGKFLQULNICQGNIGGRIVDY	243
Qy	241	ILSSONRVVRNPGERNHYHIFALLAGLEHEEREEFPYLSTPENHYLNQSGCEDVEDTSD	300
Dd	244	LL-EKNRVVRRIPGERNHYHIFALLAGLQDEEREEFPYLSTPENHYLNQSGCEDVEDTSD	302
Qy	301	QESPREVITANDWQFSKEEYREVSRLLAGILHLNIEFITAGNOVSFKTALGSGAELL	360
Dd	303	QESPREQVITAEWQFSKEEYREVSRLLAGILHLNIEFITAGGAQIPEKTAALGSGADLL	362
Qy	361	GLDPTQLDALATORSMPFRLGEEIILPLWVOQAVDSRDSLAMALVYACCEFWYIKKNSRIK	420
Dd	363	GLDPTQLDALATORSMLRGEIILPLSVQOAVDSRDSLAMALVYARCEFWYIKKINSRIK	422
Qy	421	GNEDFKSGIILDIQGFENFENVHFEQFININANEKLOEYFNKHITSLEOLEYSGELWYE	480
Dd	423	GKDFPKSGIILDIFGFENFENVHFEQFININANEKLOEYFNKHITSLEOLEYSGELWYE	482
Qy	481	DIDIMDNESCDLLEKXGLLALINESHPPOATSTLLEKSHOAHNHFFVYKRVAVN	540
Dd	483	DIDIMDNESCDLLEKXRLALINESHPPOATSTLLEKSHOAHNHFFVYKRVAVN	542
Qy	541	NFYGVYHAYGAVQYDVYRGILKXNRDTRFDLLNLRESRDFIYDLFEHYVSHNNDTIKC	600
Dd	543	NFYGVYHAYGAVQYDVYRGILKXNRDTRFDLLNLRESRDFIYDLFEHYVSHNNDTIKC	602
Qy	601	GSKRRRPVSSQPKVDSLHSLMATLSSSNPFVRCIKRMQMPQPODAVVLNOLARISG	660
Dd	603	GSKRRRPVSSQPK-DSLHSLMATLSSSNPFVRCIKRNTQMPQPODAVVLNOLARISG	661
Qy	661	MLETVIRIKAGAYARPQDFYKRYVLMRNIALPEDYRGKTSLLQYDASNSEMQK	720
Dd	662	MLETVIRIKAGAYARRPQDFYKRYVLMRNIALPDPDRGKTVLLQYDASNSEMQK	721
Qy	721	TKVPLRESLEOKLXKREEBVSHAAMVIRAHVGLFIAKQYKVLVYCVIITQKNYRAELL	780
Dd	722	TKVPLRESLEOKLEKREEBEIDRAMVIRAHVIGLVIKQYKVLVCGVTTIQKNYRAFLA	781
Qy	781	RRRFLHLKKAATVPOKOLRGOLARVYVQULLAEKREOEKKGQOEBEKKKKREEBREBR	840
Dd	782	RKKFLHLKKAATVPOKOLRGOLARVYVQULLAEKRELEKKGKREEB--KKREEBREBR	839
Qy	841	ERRBAE-LRAQOEETRRQOELALOKSEKELTRELLEKOKENQOVEIILRLKEIBDL	899
Dd	840	AQREBIDLRAQOEAETRRQOELALOKSGORAADLTRELLEKORENKQOVEIILRLKEIBDL	899
Qy	900	QRMKEQOELSTEASLOTLQERRDQELRLLEBACRAAOEPLSINPDEIDECVNIERS	959

Db	900	QRMKRGSLSTLNASLQCLQOLRBEELRLRLEDEACRAAQFELBESINPEIDECVRIENRS	955
Qy	960	LSGSEBFS-----SELASACEKCNPNFSQYP-EEVDEGEFADDDAFKDSPNPSEHG	1014
Db	960	LSVSGEISEGELSEIASASGEKPSFNFSQYPPEEBEVEDEGFADDDAFKDSPNPSEHG	1019
Qy	1015	SDQRTSGIRTSDDSEBDPYNNDVVPSPBADSUTVLAPVQDSGLJHNSGSESTYCM	1074
Db	1020	SDQRTSGIRTSDESSSEBDPYNNDVVPSPBADSUTVLAMQDSALJHNSGSESTYCM	1079
Qy	1075	PONAGDLPSPGDYVDODDYEDPAITSGSSVTEFSNSYGQSMSPDYCSVGTYNSSGAYR	1134
Db	1080	PONAGDLPSPGDYVDODDYEDPAITSGSSVTEFSNSYGQSMSPDYCSVGTYNSSGAYR	1139
Qy	1135	FSSEGAQGSFFDSEEDPDSRFDTODELSYRRDSVYSCVTLPLYFHSFLYMKGLMNSWKRR	1194
Db	1140	FSSEGAQGSFFDSEEDPDSRFDTODELSYRRDSVYSCVTLPLYFHSFLYMKGLMNSWKRR	1199
Qy	1195	WCULKDEFTFLMFRSKQEAHLKQGMHLKKGGSSTLSRPMWKRMVNLQSKLMTFENDEE	1254
Db	1200	WCULKDEFTFLMFRSKQEAHLKQGMHLKKGGSSTLSRPMWKRMVNLQSKLMTFENDEE	1259
Qy	1255	KLKGTVEVRTKEIITDNTKENGDIIMADSTPHLIASPEDAQMFVSLSQVASTDQE	1314
Db	1260	KLKGTVEVRTKEIITDNTSKENGDIIMADSTPHLIASPEDAQMFVSLSQVASTDQE	1319
Qy	1315	IOEMHDEQANPQNAVGLDVGILDSVCASDS PDRPNSFVITITANRVLHCNADTEEMHHW	1374
Db	1320	IREMHDEQANPQNAVGLDVGILDSVCASDS PDRPNSFVITITANRVLHCNADTEEMHHW	1379
Qy	1375	ITLLQRSKGDTRVSGOEFYVNGMHLKEVXNSPKMSLKLKRRMVLVTHNSLDYTKSSBKN	1434
Db	1380	ITLLQRSKGDTRVSGOEFYVNGMHLKEVXNSPKMSLKLKRRMVLVTHNSLDYTKSSBKN	1439
Qy	1435	ALKLGTLVNLSCSVPPDEKIPKETGYMNVTYGRKHCYRLYTKLNEATRMSGVIONV	1494
Db	1440	ALKLGTLVNLSCSVPPDEKIPKETGYMNVTYGRKHCYRLYTKLNEATRMSGVIONV	1499
Qy	1495	TDTRAPIDTPQOOLIODIKENCCLNSDVVEQIYKKNPILRYTHRLHPSLPLPYGDINLN	1554
Db	1500	TDTRAPIDTPQOOLIODIKENCCLNSDVVEQIYKKNPILRYTHRLHPSLPLPYGDINLN	1559
Qy	1555	LLKKQGYTTLODEAKIKFNSIQOLESMSDPPIIOGILQTHDRPLRLDELYCQLIKOTN	1614
Db	1560	LLKKQGYTTLODEAKIKFNSIQOLESMSDPPIIOGILQTHDRPLRLDELYCQLIKOTN	1619
Qy	1615	KVPHPGSGNLYSMOILTCUSCTFLPSRGLIKLYLKFHLKRIREOPGTEMEKVALFTYES	1674
Db	1620	KVPHPGSGNLYSMOILTCUSCTFLPSRGLIKLYLKFHLKRIREOPGTEMEKVALFTYES	1679
Qy	1675	LKKTCKREBFVNSRBEIEALHROEMTSYVYCHGGGSKCITINSHTTAGVEVEKLIRGLAM	1734
Db	1680	LKKTCKREBFVNSRBEIEALHROEMTSYVYCHGGGSKCITINSHTTAGVEVEKLIRGLAM	1739
Qy	1735	EDSRNMFALFEYNGHVDKALIESRTVADVLAKEFKLAATSEVGLPMKFTFKLYCFLDTD	1794
Db	1740	EDSRNMFALFEYNGHVDKALIESRTVADVLAKEFKLAATSEVGLPMKFTFKLYCFLDTD	1799
Qy	1795	NVPDSDVFEAFMEFOAHAVHGHHPARENLQVLAALRLQYLOGDVTLHAIRPLEBVY	1854
Db	1800	NVPDSDVFEAFMEFOAHAVHGHHPARENLQVLAALRLQYLOGDVTLHAIRPLEBVY	1859
Qy	1855	SLQRLKARISGSTKTFTPCERLSEKRRAPSLFEGTLRRSPRTGSVVRQKVEEQMDMMIKE	1914
Db	1860	SLQRLKARISGSTKTFTPCERLSEKRRAPSLFEGTLRRSPRTGSVVRQKVEEQMDMMIKE	1919
Qy	1915	EVSSARASIIDKMKRFOQMNOBQAMAKYMLIKEMPGYSTLFDVECKEGFPQELMLGV	1974
Db	1920	EVSSARASIIDKMKRFOQMNOBQAMAKYMLIKEMPGYSTLFDVECKEGFPQELMLGV	1979
Qy	1975	SADAVSYTKGEGAPLEVFOYEHLSLGAPLANTYKIVDERELLFETSEVVDVAKLMA	2034
Db	1980	SADAVSYTKGEGAPLEVFOYEHLSLGAPLANTYKIVDERELLFETSEVVDVAKLMA	2039



QY 2035 YISMIYKRYSTTRSSASQSSR 2057  
 DB 2040 YISMIYKRYSTTRSSASQSSR 2062

## RESULT 2

ID Q9R350 PRELIMINARY; PRT; 924 AA.  
 AC Q9R350;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Similar to myosin X (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBITaxId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC024692; AAH24692.1; -.  
 FT NON\_TER  
 SQ SEQUENCE 924 AA; 106409 MW; F4P4E9E4A1C3819B CRC64;

Query Match 44.2%; Score 4744; DB 11; Length 924;  
 Best Local Similarity 97.2%; Pred. No. 1.1e-241;  
 Matches 888; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 1134 RFSSEGAQSFEDSEDPFSDPTDELSTYRDSVYSCVTLPYFHSFLYKMGGLMNSWKR 1193  
 DB 1 RFSSEGAQSFEDSEDPFSDPTDELSTYRDSVYSCVTLPYFHSFLYKMGGLMNSWKR 60  
 QY 1194 RWCVLKDETFPLMRKQZALKQGMHLKGGSGSTLSRNNKKRWFLVLRQSKMTYFENDSE 1253  
 DB 61 RWCVLKDETFPLMRKQZALKQGMHLKGGSGSTLSRNNKKRWFLVLRQSKMTYFENDSE 120  
 QY 1254 EKLKGTVEYRTAEIIDNTKENGIDIMADRTFHLIAEPEASQMFSLVSOVHASTQO 1313  
 DB 121 EKLKGTVEYRTAEIIDNTKENGIDIMADRTFHLIAEPEASQMFSLVSOVHASTQO 180  
 QY 1314 EIOEMHDEQANPQNAVGLDVGILDSVCSDSDBRPNSFVITANFVLMCNADTPPEMH 1373  
 DB 181 EIREMHDEQANPQNAVGLDVGILDSVCSDSDBRPNSFVITANFVLMCNADTPPEMH 240  
 QY 1374 WITLLQRSKGDTRVEGOEPIVRGMLHKEVNSPKMSGLKKRPVLTNNSLDYYSSEK 1433  
 DB 241 WITLLQRSKGDTRVEGOEPIVRGMLHKEVNSPKMSGLKKRPVLTNNSLDYYSSEK 300  
 QY 1434 NALKGLTVLANSICSVVPPDEKIFKGTGVNNTVYGRKHCYRLYTKLNLNATRWSSAION 1493  
 DB 301 NALKGLTVLANSICSVVPPDEKIFKGTGVNNTVYGRKHCYRLYTKLNLNATRWSSAION 360  
 QY 1494 VTDTKAPIDPTQOOLQIDIKENCLNSDVVOQIYKRNPILETHHPLHSPILPYGDIIN 1553  
 DB 361 VTDTKAPIDPTQOOLQIDIKENCLNSDVVOQIYKRNPILETHHPLHSPILPYGDIIN 420  
 QY 1554 NLLKDKGYTTLQDEBAIKIFNSLQOLESMSDPPIIOGILCTGHDLPRLDELVCOLIKOT 1613  
 DB 421 NLLKDKGYTTLQDEBAIKIFNSLQOLESMSDPPIIOGILCTGHDLPRLDELVCOLIKOT 480  
 QY 1614 NKVPHGSGVNLVSMOITLCTPLBSRGLIKYLPFKHRIEOPGTMEKALFTYE 1673  
 DB 481 NKVPHGSGVNLVSMOITLCTPLBSRGLIKYLPFKHRIEOPGTMEKALFTYE 540  
 QY 1674 SLKTKCREFPVSRDEIEALIHROEMSTIYCHGGSGCKITINSHTAGSEVKLRGLA 1733  
 DB 541 SLKTKCREFPVSRDEIEALIHROEMSTIYCHGGSGCKITINSHTAGSEVKLRGLA 600  
 QY 1734 MEDSRNMFALFEYNGVDAKIAESRTVAVDLAKFEKLAATSEVGDLPWKYFPLKYPCLDT 1793  
 DB 601 MEDSRNMFALFEYNGVDAKIAESRTVAVDLAKFEKLAATSEVGDLPWKYFPLKYPCLDT 660

QY 1794 DNPVDSVEFAMFEQAEAVLHGHHPAEBNLQVLAALRLQYLOGDYTLHAALPPEEV 1853  
 DB 661 DSMFDSVEFAMFEQAEAVLHGHHPAEBNLQVLAALRLQYLOGDYTLHAALPPEEV 720  
 QY 1854 YSLQRLKARISOSTTFTPTCEERLEKRTSFLBGLTRRSFRGVSIVRKAYEEOMLDMWK 1913  
 DB 721 YSVQRLKARISOSTTFTPTCEERLEKRTSFLBGLTRRSFRGVSIVRKAYEEOMLDMWK 780  
 QY 1914 BEVSSARASIIDKMRKFGQMNQEOAMAKYMLIKEMPGVSTLPVECKEGGFPOELMKG 1973  
 DB 781 BEVSSARASIIDKMRKFGQMNQEOAMAKYMLIKEMPGVSTLPVECKEGGFPOELMKG 840  
 QY 1974 VSADAVSYKRGGRPLFVFOYEHILSFQAPLANTYKIVDERELLFTSSEVVDVAKLMK 2033  
 DB 841 VSADAVSYKRGGRPLFVFOYEHILSFQAPLANTYKIVDERELLFTSSEVVDVAKLMK 900  
 QY 2034 AYISMIYKRYSTTRSSASQSSR 2057  
 DB 901 AYISMIYKRYSTTRSSASQSSR 924

## RESULT 3

ID O94893 PRELIMINARY; PRT; 712 AA.  
 AC O94893;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE KIAA0799 protein (Fragment).  
 GN KIAA0799.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBITaxId=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRN;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XI.  
 RT for large proteins in vitro."  
 RL DNA Res. 5:277-286 (1998).  
 DR EMBL; AB018342; BAA34519.1; -.  
 DR InterPro; IPR000299; Band\_4.1.  
 DR InterPro; IPR000857; MYTH4.  
 DR Pfam; PF00784; MYTH4; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00295; B41; 1.  
 DR SMART; SM00139; MYTH4; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PS50057; BAND\_41\_3; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 2.  
 FT NON\_TER  
 SQ SEQUENCE 712 AA; 82011 MW; CC59A9072350B8F5 CRC64;

Query Match 30.7%; Score 3294; DB 4; Length 712;  
 Best Local Similarity 99.4%; Pred. No. 1.9e-165;  
 Matches 624; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1337 IISVCSASDSDPDRPNSFVITANRVLHCNADTPEEHHWITLLQRSKGDTRVGOEPIVRG 1396  
 DB 1 IISVCSASDSDPDRPNSFVITANRVLHCNADTPEEHHWITLLQRSKGDTRVGOEPIVRG 60  
 QY 1397 WLLHKEVNSPKMSGLKKRWFLVLTNNSLDYYSSEKVALKLGTLVNLSCSVVPPDEK 1456  
 DB 61 WLLHKEVNSPKMSGLKKRWFLVLTNNSLDYYSSEKVALKLGTLVNLSCSVVPPDEK 120  
 QY 1457 FKETGVNNTVYGRKHCYRLYTKLNLNATRWSSVITQNTVTKAPIDPTQOOLQIDIKEN 1516  
 DB 120 FKETGVNNTVYGRKHCYRLYTKLNLNATRWSSVITQNTVTKAPIDPTQOOLQIDIKEN 1516

Db 121 FKGTGWNVTYGRKCYRLYLKLNLEATWSSAIONVDTKAPIDPTQOIIQDIKENC 180  
 Qy 1517 LNSDVVEQIYKRNPIILRYTHHPLSPPLPYGGINLNLKDGYYTLLODEAIKIFNSIQ 1576  
 Db 181 LNSDVVEQIYKRNPIILRYTHHPLSPPLPYGGINLNLKDGYYTLLODEAIKIFNSIQ 240  
 Qy 1577 QLESMSDPIPIIOGIIQOTGHDLPLRLDELVCQIIKOTNKVPHGVSQNLYSWQILCLSC 1636  
 Db 241 QLESMSDPIPIIOGIIQOTGHDLPLRLDELVCQIIKOTNKVPHGVSQNLYSWQILCLSC 300  
 Qy 1637 TELPSRGILKYLKFLKRIEOPFGTEMKALFTYESLKKTCREVPREDIEALIR 1696  
 Db 301 TELPSRGILKYLKFLKRIEOPFGTEMKALFTYESLKKTCREVPREDIEALIR 360  
 Qy 1697 QEMTSYVYCHGGGCKITINSHTTAGVEVKLIRGLAMEDSRMNFALFENGVAIAIES 1756  
 Db 361 QEMTSYVYCHGGGCKITINSHTTAGVEVKLIRGLAMEDSRMNFALFENGVAIAIES 420  
 Qy 1757 RTVVADVLAKFEKLAATSEVDDLPMKRYFKLYCFELDTDNVPKDSVEPAFMEQAEHVAIH 1816  
 Db 421 RTVVADVLAKFEKLAATSEVDDLPMKRYFKLYCFELDTDNVPKDSVEPAFMEQAEHVAIH 480  
 Qy 1817 GHHPAPEBNQVLAALRLQYLOQDYLTHAIPLEEVYSIQRLKARISOSTKFTFCERL 1876  
 Db 481 GHHPAPEBNQVLAALRLQYLOQDYLTHAIPLEEVYSIQRLKARISOSTKFTFCERL 540  
 Qy 1877 EKRTSFLGTLRLSPRTGSVVRQKVEEQMLDMWIKEEVSASARASIIIDKMKFGQGNQOE 1936  
 Db 541 EKRTSFLGTLRLSPRTGSVVRQKVEEQMLDMWIKEEVSASARASIIIDKMKFGQGNQOE 600  
 Qy 1937 QAMAKYMALIKEMPGYSTLFDVECKEG 1964  
 Db 601 QAMAKYMALIKEMPGYSTLFDVECKEG 628  
 RESULT 4  
 ID Q9UIM8 PRELIMINARY; PRT; 2357 AA.  
 AC Q9UIM8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Class VII unconventional myosin.  
 GN MYO1.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE OF 254-349 FROM N.A.  
 RA MEDLINE=95023928; PubMed=7937787;  
 RA Titus M.A., Kuspa A., Loomis W.F.;  
 RT "Discovery of myosin genes by physical mapping in Dictyostelium.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20045026; PubMed=10574761;  
 RA Titus M.A.;  
 RT "A class VII unconventional myosin is required for phagocytosis",  
 RL Curr. Biol. 9:1297-1303(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Titus M.A.;  
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL, L35321; AA06035.1; -.  
 DR HSP, P08799; 1MND.  
 DR InterPro: IPR000299; Band\_4.1.  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR001609; MYTH4.  
 DR InterPro: IPR000857; MYTH4.  
 DR InterPro: IPR000159; RA\_domain.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00612; IQ\_3.

DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00784; MYTH4; 2.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000066; SH3; 1.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00295; B41; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR SMART; SM00139; MYTH4; 2.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS0057; BAND\_41\_3; 2.  
 DR PROSITE; PS00002; SH3; 1.  
 DR SH3 domain.  
 SQ SEQUENCE 2357 AA; 268281 MW; FC667F12F207BDC2 CRC64;  
 Query Match 17.9%; Score 1921; DB 5; Length 2357;  
 Best Local Similarity 25.6%; Pred. No. 28-92; Indels 769; Gaps 81;  
 Matches 652; Conservative 402; Mismatches 721;  
 Qy 62 YDMASLTLEHGGSIMVNLFORYKRNQIWTYIGSIIASVNPQIAGLYEPATMEQYSR 121  
 Db 15 VEDMTLPLTESLILNLKRYKKKEIYTYGSLVAVNPYE-ILPITYADIYSYPAK 73  
 Qy 122 HLGELPHIFAIANECYRCLMRHDNQCILKGSAGCTESTKILKFLSVISQOSLEL 181  
 Db 74 SNMLMPLHFAVSDAFTMIEGKNQSIIGESGAGCTESTKILIIQYLA----- 124  
 Qy 182 SLKETSCTYERAILSSPIMEAFGNKATVYNNSSRFQGVOLNCOQGNLOGGRIVDCI 241  
 Db 125 ARTNHSQVEQMIIVSSPLLEAFGNKATIRNNSSRFQGVOLNCOQGNLOGGRIVDCI 184  
 Qy 242 LSSQNRVVRQNGPENRYHIFYALLAGLEHEEREFEYLSTPENRYHILNOSGVEDXTISPO 301  
 Db 185 L-EKRSISQASSENRYHIFYALLAGLEHEEREFEYLSTPENRYHILNOSGVEDXTISPO 243  
 Qy 302 ESFREVITAMDVQPSKEEVRVSLAGILHGNIEFT-----AGAQVSFKTALGR 355  
 Db 244 EDFEHKYAMNVLGPEDKQFTFISVAVLHIGLKEKSEKTOGAESEVSNKDTLKI 303  
 Qy 356 SHELGLDPTQTLDTLQSMFLRGEELITPLNVOQAVDSRSLMALYACFEWVYIKI 415  
 Db 304 IAGLLSDVPVKLETCTIRHVLIRQGNFVPLKVAEADTRSLKALGVNFMVLVFT 363  
 Qy 416 NSRI-KGNEDFKSIGLIDFGFENEVNHFEQFNINVAINEKLOEYFNKIIIFLEOLEYSR 474  
 Db 364 NKRHKPQKNSFTFGLDIFGFENFKNSFEPFCINPANEKQCHFNQHIIFLEOLEYSR 423  
 Qy 475 EGLWEDIDWIDNGECLDLEKK-IGLLALINESHFPOADSTLLEKLSQHANNHRY 533  
 Db 424 EKINNSKIYVNDNOCLDLEKKRPLGILSLDEESRFPQATDLYDKLHTNHEKHPYE 483  
 Qy 534 KRVAVNNGVGVYGEVQYDVGRGLEKRRDFFRDLNLNLESRFPDYIDFEHYSSN 593  
 Db 484 KRRSKNTFVVRHVGVEVYDTQGLDKNDVSDLSLQSGSKFTIELEF--TPPE 541  
 Qy 594 NODTLKCSKRRPVTSSQFKVDSLSHLMATLSNPFVRCIKENMQMPDQFQAVVL 653  
 Db 542 EGDSDDKRKKKTTAGQCFKT-QLOSILNLSSTQPHVRIKIKNTKEPARYRELQ 600  
 Qy 654 NOLRYSGLTEVIRIRAGYAVRFPQDYKRYKVLNRNLALPEDEVGCKTSLQLYDS- 712  
 Db 601 AQLRVAWMETIRIKLGYPIRHTKEFRDRLYLDYR-ARSTDHKQTCAGILNLSGTG 659  
 Qy 713 ---NSEWQGTQVLRLEBQKLE-----KRREEVSHAM 746  
 Db 660 GLERDEWQGNKTVIRHQQYKLEBLRKLKLLKVTLLQSVWRMYRCKKRYQOQIRASK 719  
 Qy 747 VIRAHVGLFIAKORYKVLVYVITIKYRATLRRRFLHKKAAIVFQKQIRGOIARV 806  
 Db 720 IIGAAMLSHSRDQDQORAVRIKGPFKMLTYQKQFIQIINRIYQNNIRSFARRH 779  
 Qy 807 YRQLLAERKEQ-----EEKKQEEBKKKRREERER---ERBRREALRQAQEEETR 857

```

Db      780 SNNVLLKRDNRNMLEIQREKOEERNQOEKEDROEKEDKEDTDRROLOEBOQRK 839
Qy      858 OQELALOKSOKAEILTRLEKOKENKQVEILTRLEKEDLOFM-KEOQELSTLEA--- 913
Db      840 BEELBA---KREEBELKLEBKKSOUKEINOI---DELSLERMLKEQDDKONIELDDF 892
Qy      914 --SLQ-----KLOERDOELRLE- BEACRAAOEFLSINPDEIDEC 952
Db      893 VNSLEAFPEGVDDSDQPYSPNNHONYEMSPBALDKSITDLQGLKQYTRSTYKFEVDSE 952
Qy      953 VANIERISJSGSEFSELEASACEEKPNFSGQPIPEEVDGFEADDAFQDSPNPSBH 1012
Db      953 ---KPELPPGIENVLKRAFGIKRQASFLPGQPIPDVYS3POYPVDE----- 996
Qy      1013 GHSDDRTSGIRTSDDSEEDPYNN-----DTVPTSPASDVTWLPASVQDSSGLH----- 1063
Db      997 -----ADDDSDNNNTYINSNGDLPPLTSQSSDPSLPPPS---SSMDPGLEP 1041
Qy      1064 --NSSSGESTYCMQPNAGDLPSPDGDYDODDYEDGA-----ITSGSVTFSSNSYGSQW 1116
Db      1042 PPPSSSSGCTYSLP-----PWP--VFDFGMIDPLIGAPPPPSDSTSPTSATNGN-- 1091
Qy      1117 SEDYCSVGTVNSGAYRFSSEGAOSS-----FEDSEEDFSRFD--- 1156
Db      1092 -----NTPNSS-----SASASOSTNQVNPQPTVSVELPOLINDEEISLSPFYDAN 1138
Qy      1157 -----TDELSTYRSDVYSCV-----TLFYHS----- 1179
Db      1139 KNFNIKLOKQDDIFSYQSHIKSSLLVHSDABQTKVAVEFTSKVLHYNSPLVSKOP 1198
Qy      1180 -----FLYKGGIAMSWMKRWCVLKDETF-----LW----- 1205
Db      1199 ADPYSPVKILTKGLAIBS-----LRDEICYQLIKOSTNSPIQDINRVELHFTCS 1251
Qy      1206 -FRSQOALKQCMHLKGGSGSTLSRRNKK-----RHFVLROSKLMTENDSEBL 1256
Db      1252 TEPPTPKLIKYPAAYLK---TTIQSDYKSKVDSQAQSAQSYLQRFLL---NGARKOV 1303
Qy      1257 KCTVTRTAKE-----IIDNT-----KENGID 1279
Db      1304 PSVTELESJIKENRPFVRIITATDGLKGLHDSATTCQSSINDLSQRSRMVNSKENGFT 1363
Qy      1280 IIMADRTFHLIASEPEDAQWFSVLSQV---HASTDOEIOE-----MHDEQANQ 1326
Db      1364 II---ESPFGIERDLAPTDKLDVLSKVENLQATLSKIQVNFKPYFKKCLFDD---NTT 1417
Qy      1327 NAVGTLDVG-----LIDVCASDSPDRPNSFVITANRVLHGNADTPEEMHMT-- 1375
Db      1418 NNVPFTSINVNEBFYHQLFNDLFNSNYCQDQYQISIGSLKQFESSDYTDEIRAMLPBG 1477
Qy      1376 -----TILQRSKCDTRVEGOEFYVRGMLKEV---KSPYKMSLSKLKCR-----WVY 1419
Db      1478 NGRKGYTTTIDERNRPD-----DFINKYKSHGKSPEBAFKQWVQLLEKHPLANCSLVV 1531
Qy      1420 LTH--NSLDYKSSSEKNAKLGTLVNSLCSVVPPEKIFKEETGYNNVTYVGKHCYRLY 1477
Db      1532 CEHQSSSLPYKPNF-----VLALNNGINIIDPASKVLESKYSN-----OQO 1575
Qy      1478 TYLLENTVMSVAVIONVDTKAPIDPTQOLIODIKENCLM-----SDV 1521
Db      1576 QNLKSDKSVSIILENKSTLQA-FTGDVQKLVLSIKESVLYLRNNAKVARALKDYNVSDT 1634
Qy      1522 VEOIYKRNPLKRYT----- 1535
Db      1635 SLPPFGKNDITTTTFKQENKMFQGLNGKSGSPVDHVEILSDVPPQVHPVATLSLP 1694
Qy      1536 ----- 1535
Db      1695 PMSPTIWNTPPPPSISDSMSPPPOYQMLPPEPPSPVSGTKRIEIPSLGIPPPPS 1754
Qy      1536 -----HHPLHSPL--LPLPYGDINLNLKDKGYTT----- 1563

```

```

Db      1755 SSNSVSPNSPIGSPMGIPPPPTITSVHLSNSGNSITPPPLSLSTPTSTPPSLSP 1814
Qy      1564 ----- 1563
Db      1815 PNFSSLRVSMNLNDSNGDSSDDPSKRLTVSPALIGTDSQAQASTRPSFKRASTLN 1874
Qy      1564 -----LODEAIKIFNSLOQLSMSD-PIP- 1586
Db      1875 QOQATLKRAKAPVDPTAFYFNKDPKESLIEEMAKLSKAIKNFSEIMWM--MGDYP1PK 1932
Qy      1587 -----IIQGLQCHDRLPLRDELYCOLIQTQNVHPRGSVGNLYSQIILTCSCTFLP 1640
Db      1933 GQTSASLVQISIISSRIENHEIRDELYCOAYQNTNPNVESAKK--GSELVPLSITSP 1990
Qy      1641 SRGILK-YLKPHLKR---IREQFPG-----TEMKVALFTYESLKTKECFVPSR 1687
Db      1991 SDSLQPPMEQMSNMNIOSSSPQASLAVICIEKLESH1PISQO-----RMGSA 2044
Qy      1688 DEIEALHROEWTSVYCHGGGSCKI-----TNSHTTAGEVVEKILRGLAM-ED 1736
Db      2045 TEI-----QSFSSYLNENGDISTCKIRPIDOSTKLAKINTYTTIREINDVCRQYGISQ 2098
Qy      1737 SRNMFALFEYN--GHVDAIISRTVADVLAKFEKLATSVEGDLPMKFPYKLYCFDLD 1794
Db      2099 SIKMFISAVNETAGISRVSEBTDMLYDLARWEQ---SEKGR---FYQVRRRPFLD 2151
Qy      1795 NVPK-----DSVEFAFMEQAEAVIHG--HHPAPEENTQVLAALRLQYLGQDYTL 1843
Db      2152 DVNKLIDHEMLMTDDICFELTYCQIRDEMKGLYTNNEKDSIIMAILQILYPNOS- 2210
Qy      1844 HAAIPPLEEYVSLQRLKARISQSTKTPPCERLEKRRTSFLEGLRRSPFGSVVRQVE 1903
Db      2211 -----KLVLTFGEVVRQYLP 2224
Qy      1904 BEQMDMNIKEEVSARASIIDKMRKFOGMOEQAMAKYM--ALIKXMPGSGTLFVBE 1960
Db      2225 DOIINSONIKNWISIESOI-----FBLVSQTPREYLLKMTINLIGSISPLFGCTLPNQ 2278
Qy      1961 CKEGFPQELMWIGVADAVSYV--KRGGRPLEVQYEHILSFGAPLANTYKIYVDER-- 2016
Db      2279 OKEN--PRKANLAINKGVISIFDPHTKESKXFWTQISINAVF---TIDTFCIMGNLMK 2333
Qy      2017 --ELLFETSEVVDVAKMKAYIS 2037
Db      2334 PIKQTFITDEHSSIASVQFYSS 2356

RESULT 5
Q9V326
ID Q9V326 PRELIMINARY; PRT; 2167 AA.
AC Q9V326;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE CK gene product.
GN CK OR CG7595.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Change M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baxendale J., Bayraktaroglu L., Beasley B.M.,

```

RA Besnor K.Y., Benos P.V., Berman B.P., Bhandari D., Botshakov S.,  
RA Beckson D., Borchan M.R., Bouck J., Brokstein S., Brottier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fodler C., Gabriellian A.E., Gair N.S., Galbati P.W., Glaeser K.,  
RA Gloder A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J., Ketchum K.A.,  
RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kvit Z., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mikhlov G., Milebina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,  
RA Palazzolo M., Plattman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splyer E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Ventel E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Wadgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye Y., Yeh R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RC  
RX MEDLINE=99403001; PubMed=10471707;  
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,  
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whiteaw K.,  
RA Cenhiter S., Rubin G.M.;  
RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
RL *Drosophila melanogaster*: the *Adh* region.";  
RN Genetics 153:179-219(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RC  
RA Cenhiter S.E., Aghayani A., Arcaina T.T., Baxter E., Blazej R.G.,  
RA Butenoff C., Chambe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,  
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,  
RA Lewis S., Li P., Lomocan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,  
RA Nixon K., Paclet J.M., Park S., Pfeiffer B., Poon L., Sequerra A.,  
RA Sehl H., Solt E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,  
RA Zieran L.L., Rubin G.M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL: AEO03466; AAF53435.1; -;  
DR EMBL: AEO03412; AAF4915.1; -;  
DR HSSP: P08799; 1MND;  
DR FLYBase; FBgn0000317; ck;  
DR InterPro; IPR000299; Band\_4.1;  
DR InterPro; IPR000048; IQ region;  
DR InterPro; IPR001609; myosin\_head;  
DR InterPro; IPR000857; MYTH4;  
DR InterPro; IPR000567; SBP\_dac\_1;  
DR InterPro; IPR001452; SH3;  
DR Pfam; PF00612; IQ; 4;  
DR Pfam; PF00063; myosin\_head; 1;  
DR Pfam; PF00784; MYTH4; 2;  
DR Pfam; PF00018; SH3; 1;  
DR PRINTS; P000193; MYOSINHEAVY;  
DR ProDom; PD000035; myosin\_head; 1;  
DR SMART; SM00295; B4; 2;  
DR SMART; SM00015; IQ; 3;

Query	Match	Score	Length	DB	DB 5	Length	DB 5
DR SMART; SM00242; MYSC; 1.	17.2%;	1843.5;	2167;				
DR SMART; SM00139; MYTH; 2.	26.4%;	1843.5;	2167;				
DR SMART; SM00326; SH3; 1.	Pred. No. 2.2e-88;						
DR PROSITE; PS50057; BAND_41_3; 2.	Matches 629; Conservative 381; Mismatches 797; Indels 573; Gaps 88						
DR PROSITE; PS50056; IQ; 1.							
DR PROSITE; PS01037; SBP BACTERIAL_1; UNKNOWN_1.							
DR PROSITE; PS50002; SH3; 1.							
KW SH3 domain.							
SO SEQUENCE	2167 AA; 250307 MW; 3C57E34ADD089A42 CRC64;						
Query Match	17.2%;	Score 1843.5;	DB 5;	Length 2167;			
Best Local Similarity	26.4%;	Pred. No. 2.2e-88;					
Matches 629; Conservative 381; Mismatches 797; Indels 573; Gaps 88							
4	QGRVWLR-ENGQHPSTVNS---CAEG-IVPRTYGGVFTKOSTITHOKVTANPT 57						
6	RGDIYIEASGREEDVAIGARVYSAERRIQVADDGDEWLAPE-----RIKAMHAS 60						
58	NBEGVDMASTLHGGSIMYNLFQRYKRNQIMWTYIGSIILASVNPYQPIAGLYEPATMEQ 117						
61	SVQGVEDMISLDLHEAGILRNILIRYKEMIIYTGSIILAVNPYQ-ILPIYTGDIQIKL 119						
118	YSRHLGELPHIPIALNECYRCLMKRDNOCILIKESAGKTESFKLLKELSVSQ 177						
120	YKERKIGELPHIPIAIGNAAVAMKRYKODCVIISGSGKTESFKLLQYLAIS-- 177						
178	STLSLKEKTSCEVEAIILESSPIMEAFGNATVYNNNSPFGKVFQNLICQKNIOGRI 237						
178	-----GCHSMIEQIILEAPDLEAPGNATINDNSRKGKIDIHFGANVIGSAKI 230						
238	VDCTLSQNRVVRNPGERNHIFPALLAGLEHEEREPEYLTPEHYHLYNOSGCVEDKT 297						
231	EQYTL- EKSRIVSQSHSERNHVFCYIAGISABEKRLDGMADYKYLGTGNSITCEG 289						
298	ISQGESFRVITAMVMOFSKEVEYRVLLAGILHGNIEFTVA-----GGAQVSFKTA 352						
290	RDDAAEFSSIRAMKYLFPDSEIWEIKLALHHCNIIKKYKATVDINDATEIPEHIN 349						
353	LGSAELLGLDPTQDALTQSMFLNGEILITPLYNQOAVDSRSLAMALVACCEEWI 412						
350	VERVAALLGLPIQPLIDALTRTLFAHEGTAVSTLSRQSDVDADAIVKGIYRMFHIY 409						
413	KKIKNSIKKNEDPK-----SIGLIDFGFENEVNHFPQFINVYNEKLOEYFNKIF 465						
410	RKNTNVI-----FKRGTSRNALIGVLDIFGENPDQNSFEQFCINYNENMLQGFQIHF 464						
466	STEOLEYSGREGVEMEDIMDINGECLDIE-KKLGALALINESHPQATDSTLLEKHS 524						
465	KLEQEEYNEHALNMGHIEFVNDQALDIALKQINMALIDEBARPFKGDQMTLAKHK 524						
525	QPAHHFVYKPVAVN-NFGVKHAGEVQYVGRGILEKNRDTFPDDILNLRBSRPFY 583						
525	THQSHKNVYLPKPSDINTSFGLNHFAGVVFYDTRGFLDKNRDTPSLDLHVSQTYNFKLR 584						
584	DLFEHVSRRNODTLCKSKHRR--PTVSQFQKDSLSLSMAATLSSNPFVCIKRNMQ 641						
585	QIFA-----QD-IEKAEIRKRPFLISTQPR-KSLDALMKTLSQCPFFICIRNEL 635						
642	KMPDQFQAVVNLQRLYSGLTETVIRKAGAVARRPFQDYKRYKVLNRNALPEDVRGK 701						
636	KRPMPEDRLCRORLYSGMMETIRIRAGYPIRHGREFEYERYFLIP--GVPAHRTD 653						
702	C---TSILDYDANSSEWOLCKTYVLEPSLEQLEKREBEVSHAMVIRAHVLEGLAR 758						
694	QQAATSRICAVVLGKSDYQLGHTKFLDADHDLLEBRD-----VLR 738						
759	KQYKVLVCVVIQKQYRAFLRRRFLTKKAAIVFQKQGLARVYRQLAEKREOE 818						
739	K-----ILLQNSIGWYRRRFLRLRAAALTIVQRFMKGVQAKRYNNM----- 782						
819	EKKQEEBEKKKREBEEREREREAELAQ-----QEEETRKQOELEAQ 865						
783	-----RVGWMQLQALISRYLSHFRLRHGHIVGLQAHNAGYLVREYGHKMAVYIKQ 836						

```

Qy 866 KQKAEALTRLEKOK--ENKOVEELIRLEKEIEDLQRMKEQOELSTLEASLOKQJERRDQ 924
Db 837 SHVRNRIAMRRYRKLRLEHKQFAEVLQAK--LEE-----QELHRRGK 878
Qy 925 ELRRLLEECRAAOEFLBSLNDEIDECVRNIERSISGSEFSSSELAESACEKENFNFS 984
Db 879 HAREIAEQHYR---DLHELREIREIOLENNRR---VEVMNNTINDAARQ----- 924
Qy 985 QYPEREEVDEG-----FEADDAFPKDSNPSEHGHSDQRTGIRTSDDSSBEDPYMNT 1038
Db 925 ---BEPVDDGKLVEMFPFLPDSSSDAPTP---HGKRETSVF--NDLPHANVNODDI 974
Qy 1039 VVPTSPASDSTVLLAPSVODSGLSNSSGSESTYC-----MPQN----- 1077
Db 975 IAPIHISEDEEDLSEFKFQKPAATYFQGVNNOYAKKALKHPLLPHTOGDOLAQAALMI 1034
Qy 1078 ---AGDLPSPDGDYDODDYEDGATISGSSVT-----FSNYSYGOWSPD--- 1119
Db 1035 TILRFTGDMPEPK---YHTMDRMDTTSVMSKVATIGRNPIRSKEFOEAQLMGLDPAFL 1091
Qy 1120 ---YRCSVGTNSSGAYRFSEGAQSFDESEDEDS-----RPTDDE 1160
Db 1092 KQKPSIRKLVSLTLKRKQKLG---EDVARRIQDDEVTADSYQSMQSRPTSULEK 1145
Qy 1161 LSY-----RDSVYS--CVTLPYFHSFLYMKGLMSWKRWM-----CVLKD 1200
Db 1146 LHPITIGHGLRABLREIYCOJCKOL-----TNPLK3SIARMIILSLCVCGFAPS 1197
Qy 1201 EYFL--WFRS-----KQBALKQGLMHKKGGSGSTLSRNNKKRPFVLRQSK-- 1244
Db 1198 EKFVNYLRAFIREGPGYAPYCEERLKRTFNN-----GTRNOPSPMLEIQATSKK 1248
Qy 1245 ---LMYENNSEBKLKGTVEVRTAKEIINDNT-----KENGIDIIIMADRTFHIIAS 1293
Db 1249 PMLPTTFMDGNTKTLADS--ATTARELCNQLSDKISLKDOGFSLYA--LEFDVSSL 1304
Qy 1294 PBDASQFVSLSQVNA--STDQBIQ-----MHDEQAMP--QNAV 1329
Db 1305 GSGGDHVMVAIQCEQYAKEQOGBENABWRLEFRKEIFAPMHEPRHDQVATNLIIQOVV 1364
Qy 1330 GTLDVGLIDSVCASDSDPRPNSEVITIANRVLHCNADPEIMHMHTLLQRSKQDTRVGG 1389
Db 1365 RGCVKFG--EYRC-----DKSEBDMAMIAOOYF--IBYSTDMEMERLFTLLPNFIPDCLSG 1416
Qy 1390 QSEIYRGW-----LHKEVKSPPKMSLKLK-----RMFVLTNHSLODYKSEKN 1434
Db 1417 VKALERMALVLQAVYKSYVVKDIAPIKIKEDIVSAKIKWPLLSRFEAYRNSGPN 1476
Qy 1435 ALKGLTIVLNSLCS--VPPDEKIFKETGYMNTVYGRKHCVRLYTKLNEAT----- 1485
Db 1477 LRKNDVITAVNTGVVVDQBOVLELSEFPETIANVSKQKINNVFQTSLSLVRESEPT 1536
Qy 1486 ---RMSVSIQVNTDTKADIDPTJ-----QOLIODIKENCIN 1518
Db 1537 FQSPNADIRDLVYVFLDGLKKRSKVIYALQDYRABSDGTSFLSPFKGLIILEDSCGE 1596
Qy 1519 SDV-----VEQIKRNPILRYTHHPLSLPLPYGDJ-----NINLAK 1557
Db 1597 SVLNNGMCIGRCDSRQSDGDPRAETVYV--PFLSKRPDIDIALFNIIEAHGRRLSMS 1654
Qy 1558 DKGYT-----TLQDEAIKIFNSLOQLSMS-----DPI--PIIQ- 1589
Db 1655 NGGAVPRDRPHLMYALDHR--LRPKRTMSKTLTLLSGKSELMRYGRDPIKAPLKK 1713
Qy 1590 ---GILQTHDL--RP-----LRDELVCOLIK 1611
Db 1714 LOSKEFAEACPAFAAILKYMDDLSPKPRMGENEITDHFQDPLKGBILIRDEIYCOLMK 1773
Qy 1612 QTNKVPHPSPVGNLYSMOILTCLSCFLBSRGILKYLKHLKXIRQOPFOTBEKAKALT 1671
Db 1774 QL--TDNRNRMBEERGEMLMWATGLFACSGGLKELLLFLTRRRHPISODSMHR----- 1826

```

```

Qy 1672 YESLKT--KREFVPSRDEIALIHR--QEWTSVYCHGGGSKITINSHTTAGEVVEK 1727
Db 1827 ---LQKTIHRGQRKYPHQVEVEAIOHKTQOIFHKRYVPPDPTDEAFVDSSTRADFCNN 1883
Qy 1728 LIRGLAMEDSRMPL-----FEYNHVDKAYSRIYVADVAKFETL 1770
Db 1884 ISQRLSLSTSEG--FSLFKVIADKVISVEGDFPFVHLLTDWIKKARPIRD----- 1934
Qy 1771 AATSEVGLPMKFYKLYCF--LQTDNVP--KD--SVEFAFMFEQOAEVYIGHHPAPEML 1826
Db 1935 ---GANP--QFTYQVFPMKKLTNTVPGADRNDLFIHQBELPKLLRGTHKRSBEA 1987
Qy 1827 QVLAALRLQYLOGDYTLHAIPLEEVYSLOLRKARISQSTKTPPCRSLRKRTSPLEG 1886
Db 1988 AKTALVFRVRGE-----NKQELQ--IQMLRELIPSOIMKIQSTS----- 2028
Qy 1887 TLRRSFRSGSVVRQVREBQMLDMYIKESVARSASIIDKRRKFGQNOEQOMATYMALI 1946
Db 2029 ---EW-----KRSIVASYNQDGMTSEDAKVAFLKIV 2057
Qy 1947 KEMPGVSTLPDV--ECKEGFPQELMLGVSADAVYVRKGEGRPLEVQYEHILSFGAPL 2005
Db 2058 YAMPFTGSAFPEVKQTTBPNTPEMLLIANKHVSILHPYTDILVTPFTIRISWSS-- 2115
Qy 2006 ANTY-----KIYDERELLFETSEVVDVAKLMKAYISMIV 2040
Db 2116 GNTYFPMHTIGNLVRGSKLLCETSLGYKMDLLTSYISLML 2155

```

# RESULT 6

```

ID Q9DGG8 PRELIMINARY; PRT; 2179 AA.
AC Q9DGG8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myosin VIIA.
GN MYO7A.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUBERINGEN;
RX MEDLINE=20414642; PubMed=10958658;
RA Ernest S., Rauch G.U., Hafliger P., Geisler R., Petit C., Nicolson T.,
RT "Marterin is defective in myosin VIIA: a zebrafish model for human
hereditary deafness.";
RL Hum. Mol. Genet. 9:2189-2196 (2000).
DR EMBL; AJ404002; CAC05419.1; -.
DR HSSP; P08799; 1MDN.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00612; IQ; 5.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MYTH4; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B41; 2.
DR SMART; SM00015; IQ; 4.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00139; MYTH4; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50057; BAND_41_3; 2.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50002; SH3; 1.

```

SQ SEQUENCE 2179 AA; 251655 MM; FF440D1EFD588F76 CRC64;  
 Query Match 17.1%; Score 1834.5; DB 13; Length 2179;  
 Best Local Similarity 25.7%; Pred. No. 6.5e-88;  
 Matches 624; Conservative 389; Mismatches 780; Indels 639; Gaps 83;

```

QY 4 OCTRWVL-RENGOFSTVNSCAEIVFRTDYGOVFTYKOSTITTH-----OKYTAH 55
DB 6 OGVOYVLDLKTGHEFEVPI-----GAVVKLCDSGQIQVLDEGNEHMSIPNANTNKKPM 60
QY 56 PTNEBEGVDMASTELTHGSSIMTNLPORYKNOIWTYIGSLASVNPYQFIAGLYEPATM 115
DB 61 PTSHGEDMIRIGDLNEAGILRNLLIRYEHLYITYTGSILVAVPYQ-LLEPIYADQI 119
QY 116 EGYSRHILGELPHIFALANECYRCCLKRHNDQCILIKGSSGAGKTESTLILKPLSVIS 175
DB 120 RLVTNKKIGEMPHIFALADNCTFNNORNNKQOCIISSGSGAGKTESTLILQFALAS 179
QY 176 OQSLBELSLKEXTSCEVERALLESSPIMEAFGNACTVNNNSRPGKFVOLNICOKNIOG 235
DB 180 GQH-----SWIEQVLEANPILAEAFGNACTIRINDSSRFQKIDHFNKRGALIEGA 230
QY 236 RIYDCLLSQNRVVRQNGERNYHIFALLAGLEHEBEREFYISTEPNHYLNQSCVED 295
DB 231 KLEQYLL-EKSRVCROARDERNHYIFCYMLKGMTPOKQGLSKATDYTLTIGACTYC 289
QY 296 KTISDOESPREVITAMDVMOFSKEVEVRSLLAGILHLGNIE-----FITAGAGVSPK 350
DB 290 DGDHDDQKEYSINISAKVLMFTDKEKNEISKLALILHMGNLXYEARTYNLDACEVRC 349
QY 351 TALGRSAELLGDPLOTDLTQSRMFLNGEELLTPLVQOAVDSRDSLAMALYACCFEW 410
DB 350 SALTAAVLEVLVDLKDLMCLTSRTIITRGETVSTPLISIQALDVDAFVKGIGYGLFW 409
QY 411 VIKKISRIKGNDF-----KSGILIDIGFENFENYHNEQONINAYNEKLOEYENKAI 464
DB 410 IYKINAAIYKPPSELKAVRKSIGLDLFGFENFVNSPEOLCINFAMENILOQFVRH 469
QY 465 FSLQLEYSREGVWEDIDWIDNGBECLDLE-KKLGLALINESHPOATDSTLEK 523
DB 470 FKLEQERYENLENNMOHIEFTDNOALDMAIKPMNISLIDESKFPKSTDTTMLANK 529
QY 524 SOHANNHFYVKRP-VAVNNFGVKGAYGEVOYDVRGILKKNRDTFRDDLNLLESRFDEI 582
DB 530 SQKHLNTNYIPKQVTEYETGFIQHPAGVYVYETRGLEKKRDTLHGDIIDLVHSSKXKI 589
QY 583 YDLFENHSSNNODTLKCGSKHR--PTVSSQFKVDSLSHMTLSSNPFRCIKPM 640
DB 590 KQIF-----QADVAMGAEIKRSPTLSSQFK-RSLLEMLRTLSVCOFFVRICKPME 640
QY 641 OKRPDQFQAVLNQLRYSGLMETVRIRKAGYAVRRPFODFYKRYKVLAMNLA---LPED 697
DB 641 YKRPMLFRELCTVROLRYSGMTETIRIRAGYPIRTTFVAVRVRYKVLAMGVPRAYKQED 700
QY 698 VRGKCTSLQLYDANSSEWOLGKTIVFLRESLEQKLEKRESEVSHAANVIRAVLGFLA 757
DB 701 LRGTGCRIRAEAVLGRDDDMQMGKTKIFLKDHDMLLEIRDKAITDKVILIQVAVGFKD 760
QY 758 RKQYKRVLYCVVILQ-----KXY----- 775
DB 761 RSNFLMKKSAMLIQKTRGWYCYCRKNYGANRGPSRLQALYRSKLYQTVHVARQIMLF 820
QY 776 ----RAFLRRRFLHLKKAIVFOKOLGOIARVYROLLAE--KREOEKKKOESEK- 828
DB 821 QGRGRLVRRARFRHRLMAVITTOATTRGMIRARLTKRLGCEYRRRLIEAKKFLABEOKL 880
QY 829 ----KKREBERERERREAEI-----RAOEETRKOQOEALQSKOEKA--- 871
DB 881 RNQMSARKAKEBAEKHQRERLAQAREDAERKEKQEARKMVEMLDQMEKARQEVVNS 940
QY 872 ----ELTRELEKQENKQVEILLRL-EKELEDQORMK 903
DB 941 DNVDMKFGFLGTNSPFGOGQAPAGFEDLERTHRELEVDDLESILPLPEDDLEDLSBYK 1000
  
```

```

QY 904 -----EQEELS----- 909
DB 1001 FAKFATYRQGSTHTYMRRLKQPLRFHEDGDDQALAAVITVLRFGMDLPERKYHTA 1060
QY 910 -----LTEASIQLOERRDQELR-----RLEBACRAAOEFLSLNFDIED 950
DB 1061 ISDGSEKIPVMTKIYETTELKTKYRELQALQEGEGENTNIESHKSSVRRHKLVLTKKKS 1120
QY 951 ECVNRITERLSGSEFSSSLASACEKKNFNS-----QPYREEVDSGEFAD 999
DB 1121 KTBEEVTKRLNG-EYVH-GNSMLEDRPTSLKELHPIITGNGILRPGLRDI---YCOI 1175
QY 1000 DDAFKDSPNPSERH-----HSDOR----- 1018
DB 1176 CQOLNPNKSSHAGWILMSLCVGCFAPESEKFKVYKLRNFISEGPPGVAPEEERLRRTF 1235
QY 1019 TSGIRTSDDS-----SEEDPYM-----NDTVVPTSPSADSVLLAPSVQDSGLHN 1064
DB 1236 VNGTRTQPSMLELQATKSKKPIMLPVTFMDGTTKLTLDSATYAKELCNALSDKISLQD 1295
QY 1065 -----SSGSESTYCMQNGDLRSPDGDYVDQDDVEDGALTSGSSVTFEN 1110
DB 1296 RGFSLIYALFDKVSLSGSGN---DHVMDAVSQCEQYAKKEGQAOERNA---PWRLEFRK 1348
QY 1111 SYGQWSPDYRCSTVGT---YN-----SSGAYRFSSEG-AQSSFESEDPSPRPDTDBE 1160
DB 1349 EIFTFPHDPAEDAVATNLIYQQIVRGVYKGEYRCRQEDLAELASQOYVYDGSSELVERL 1408
QY 1161 L---SYRRD-SVYSCVTLPRHSLY-----MKGLMNSWKRRWCYL 1198
DB 1409 LELISYIPDRDISAKTYERWAQFIAMAKKGVYTOKKVDQKYEEVLEDFARXKMPLL 1468
QY 1199 KQETFLMFRSKQBALKQGLHKKGGSSSTLSRNKKRMFVLROS---KLWYFENDSE 1254
DB 1469 FRFYEAFKFGSPSLPK-----NDYIVAVNMTGVFVDEQEOVLLELSPETLAAS 1519
QY 1255 KLGK-----TVEVTAKEIINDTTKENGIDILIMDRFPHLAESEPDASQWFSVLSQVH 1308
DB 1520 SSKGGLQAGSTLNTATIKADEFTFTSNAEDIRDLVTF--LEGLRKSKRVVAL- 1572
QY 1309 ASTDQIEQMHDEQANPONAVGTLVGLIDVSCASDSPEPRNSFVITANRVLHCNADTP 1368
DB 1573 -----QONP-----SPAADS----- 1583
QY 1369 EEMHHITLQRSKQDTRY---EGOEPIVRMWLKEVNSPKMSLKKRW----- 1417
DB 1584 -----TFLSLKGDVLVDDTGBQVNTSGMAHG-----TNDRTKQGRDPADCV 1628
QY 1418 FVLTHNSLDYKXSEKNALKGLTVLNSLCSVVPPEDEKIFKETGYMNVTVYGRKHQYRLY 1477
DB 1629 YVLP-----TVVRPRHDV---VAVTMTPRDQROESLR- 1658
QY 1478 TKLNEATRWSVIONVDTKAPIDTPQO-----LIQIKENCLSNDVDEQIYKRNPL 1532
DB 1659 -----SHVALTEETERVVRPYLLEESYDYFRPRPRTLSRWITQNRGDKLM 1707
QY 1533 RYTHHRPLSLPLPYGQINLNLKDKGYTTLQDAIKIFNSL-----QQLS 1580
DB 1708 CCTREPIKQALK------KCGHEBSQECNAFIVMKMGYDPSKRTSRVNE 1756
QY 1581 MSDPIPIQGILOTGHDRLPLRDELYCOLIKO--TNKVRPHSPSVGNLSMQILTCLSCTF 1638
DB 1757 LRD--QIFEGALKA-----EPLADELYCOLIKQLTENHIIKYSEKG---WELMLCVGLF 1806
QY 1639 LPSRGLIKYLKHLKIRBQFQTEMEKXALFTYSLLK--TKCREVPSPDEIBALIH 1695
DB 1807 PPSNVLPHVQFRLOS-KGHP-----LALDQMRLOQALRNGSRKYRPHLVEBALOH 1859
QY 1696 R-QEMTSTYVCHGGSCKTITNSHTAGB---VVEKILRGLAMDSRMFA----- 1742
DB 1860 KTTQIFHKVYFPDDTDEAEVBSSTKADFCINTISSRL--LTPBGFSLFVKISDKVIS 1917
  
```

QY 1743 -----LREYNGHVDKALISRTVADVAKFEKTAATSEVGDLPKFTF--KLYCFLLDT 1794  
DB 1918 VEEGDFPDVHLLDWMKKARPAKXGI-----VPSLTYQVFFMKXLM---TT 1962  
QY 1795 NVP-KDS-VEFAFMPEQAEHVAIGHHPAEEHLQVLAAL--RLQVLOGDYTLHAIPPL 1850  
DB 1963 TVPGKUSPADSLFHHYQELPKYLRGKRSREBFOLGALYKVKF-BDDKSHFPSPHPM 2021  
QY 1851 BEVYSIOLKARISOSTKTPTCEERLEKRTSFLEGTLLRSFRTGSVVRQKVEEQMLDM 1910  
DB 2022 -----LKEMIPQ-----DLINQLSPDD----- 2038  
QY 1911 WKEBVSARASIIIDKRRKFOGMMQOAMAKTALIKEMFGYGTLPDV-ECKEGGFPQE 1969  
DB 2039 W-----KRSIVAVFNNHAGKSRBEAKLMFLKIIFFKWFPGSAFFEVQTTPEHPPEI 2090  
QY 1970 LMLGSADAVSYRKGEGRPLEVFOYEHLISFGAPLANTKYIV-----VDEBELLEPTE 2024  
DB 2091 LLIATNKGVSLIDPRNKDILTTYPTKLSNMS--GNTYFHTTGNLVQSKSLCETSL 2148  
QY 2025 VVDVAKMKAYISMIVKRYKRYSTTRASQSGS 2056  
DB 2149 GYKMDLLTYSIQMLT---TWSKQNSRSGS 2177  
RESULT 7  
QY 09DGG9 PRELIMINARY; PRT; 2179 AA.  
AC 09DGG9;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Myosin VIIA.  
GN MYO7A.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxId=7955;  
RN [1]  
RP SEQUENCE FROM H.A.  
RC STRAIN=GD5;  
RX MEDLINE=20414642; PubMed=10958658;  
RA Ernest S., Rauch G.J., Hafliger P., Gelsler R., Petit C., Nicolson T.;  
RT "Mariner is defective in myosin VIIA: a zebrafish model for human  
hereditary deafness";  
RL Hum. Mol. Genet. 9:2189-2196(2000).  
DR EMBL; AJ404001; CAC05418.1; -;  
DR HSSP; P08799; 1MDN.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR001609; myosin\_head.  
DR InterPro; IPR000857; myosin\_head.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00612; Band\_41; 1.  
DR Pfam; PF00612; IQ\_5.  
DR Pfam; PF00063; myosin\_head; 1.  
DR Pfam; PF00784; MYTH4\_2.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000355; myosin\_head; 1.  
DR SMART; SM00295; B41; 2.  
DR SMART; SM00015; IQ\_4.  
DR SMART; SM00242; MYSC; 1.  
DR SMART; SM00139; MYTH4; 2.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50057; BAND\_41\_3; 2.  
DR PROSITE; PS50056; IQ\_2.  
DR PROSITE; PS50002; SH3; 1.  
SQ SEQUENCE 2179 AA; 251632 MW; F53ABB5128DBBFH CRC64;

Matches 624; Conservative 386; Mismatches 784; Indels 637; Gaps 82;  
QY 4 OSTRVWL-RENGQHPSTVNSCAEGIVFRTDYGQVFTYKOSTIRH-----OKYTAHH 55  
DB 6 QGDYVWLDLKTGHEEVYI-----GAVVKLDSGGIQLVLDDEGNHMSIPQATNIKPH 60  
QY 56 PTNEGGVDDMASLTELHGGSIMYNI-PQYKRNQIVTYGSIILASVNPQIAGLYEPATM 115  
DB 61 PPSIGVEDMIRLGDINBAGILRNLLIRYREHLIYTYGSIILAVNPQI-LPIYTAQDI 119  
QY 116 EGYSRRHGELPRPHIATANECEYRCLMRHNOCLLIGESAGKTESKILKELSVIS 175  
DB 120 RLTYNKKIIGEMPHIATADNCFYMNQRNNQCCIIIGESAGKTESKILLOFLAALS 179  
QY 176 QOSLSLSKEKTSCEYRAILLESSPYMEAFGNKATYVNNNSSPFGVOLNICOXNIQCG 235  
DB 180 GQH-----SWIEQVLEANPILBEAFGNKATIRANDSSPFGYIDIHFKRAIEBA 230  
QY 236 RIVDCITSSQNRVVRQNGERNYHI FYALLAGLEHEBEFFYSTPENHYILNOSGVVD 295  
DB 221 KIEQYLL-EKSRVCRQARDERYNHI FYCWLKQMTDQKQGLSKATDYTYLTIGNCTVC 289  
QY 236 KTIQOESFREVITMDVMQFSKEVREVSRLAIGLHGNIE-----PITGAGQVSK 350  
DB 230 DGRDQKEYSNIRSAMKVMFTDKENWEISKLLAAILHMGILRYEARITYDNLDAEYVRC 349  
QY 351 TALGRSAELLGADPQTLDTALTORSMPLRGEIILPTLVQVQAVDSRCSALMALVACCFEM 410  
DB 350 SALTAAVLLEVDLADLWNCILTSRTIIRGEVSTFPLEQALDVRDAFVKIYGRLEFW 409  
QY 411 VIKKINSRIKGNED-----KSIGILDI FGFEFNVHFEQFNINYANEKLOEYFNKII 464  
DB 410 IVEKINAAIYKPPSELKAVNRISGLDIFGFENFMVNSFELCFNFANENLQGFVRRIV 469  
QY 465 FSLDELVSREGLVEMEDIMDNGECULIE-KYGLALLINESHFPQATSTLLEKHA 523  
DB 470 FTLDEBEVNLNINQHIEFDNDQALDMIAKPNIIISLIDESKFPKGTDTTMINKN 529  
QY 524 SOHANNHFFVVKPR-AVANNFGVGHYAGEVOYDVRGILEKRDTPFDLLNLRESRPDI 582  
DB 530 SOHKANTYIIPKNTYETQFGIQHPAGVYIETRFLEKRDITLGTIIQLVHSSKPKFI 589  
QY 583 YDLFEHVSRRNQDTLKCGSKRR--PTVSSQFKVDSLHSLMATLSSNPFVRCIKPM 640  
DB 590 KQIF-----QADVANGAETRRKRSPTLSSQFK-RSLELMRTLSVCPFFVRCIKPM 640  
QY 641 QMPEQFOQAVVLANLDRISGMLETYRIRKAGYAVRPPQDPYKRYKVTMLRNA---LPED 697  
DB 641 YKKPMLFRELCEVRLARYSGMMETIRIRAGYPIRYTFVEFDRYVILPQGVKPAVKQED 700  
QY 698 VSGKCTSLQLYDANSNMQGKTKVPLRESLEQLEKREBEVSHAMVIRAHYLGFLA 757  
DB 701 LRGTCQRIAEAVLGRDDQMGKTKIPLKDDHMLLEIRDAITDKVILLQKVVRGFKD 760  
QY 758 RKQYRKVLVCVILQ-----KNY----- 775  
DB 761 RSNPLMKKMSMLIKTKRGYYCRKNYGAMGCFRLQALYRSKLYOTYHVAHQRIMLF 820  
QY 776 ---DAFLRRRFLHLKKAALVFOKQRLQRLARRVYQRLAE--KREDEKKKQDEEEK- 828  
DB 821 QGRCRGFLVRAFRRLNAVITIQAYTRGMIALRLLYKLEKGYRRLLEAKRLLEBQTL 880  
QY 829 -----KKKEEEREREEREREAE-----RAQGEETRKQQLLEALQSKXA----- 871  
DB 881 RNQMSARAKAEAEKHQERLQRLARBEAEKKEKQARRKMEMLDQNEKARQBPVDS 940  
QY 872 -----ELTLELEKQENKQVEEILFL-EKEIELOQRMK 903  
DB 941 DMVDKMPFLGTTNSFPQOBQAPAGFEDLERTHRELEVEDLDESLFLPEDDLBLSSEK 1000  
QY 904 -----BOQELS----- 909  
DB 1001 FAKFSATYFOGTSTHTYRRPLKQPLFHEDEGQLAALAVITLVRPMGDLRBPKYHTA 1060



QY 910 -----LTFEASLOKLOERRDOELR-----RLEEFACRAOEFLESINFEID 950  
 Db 1061 ISDSEKIPVMTIYELGKTKYRELQALQGEENTHIESHKSSVRLHVLTLTKKS 1120  
 QY 951 ECVRNIRLSIGSGSFSSSBLASACEKNFNFS-----QPYEEVDEGFEAD 999  
 Db 1121 KITBEVTKRLNDG-EYTVH-GNSMLBDRPTSNEKLHFIINGILRGLADEI---YCOI 1175  
 QY 1000 DDAFKOSPNSSEB-----HSPOR----- 1018  
 Db 1176 CQJLONPSSKSHARGWILMSLCVGCAPASEKVKYLRNFISSGPGVAPICERLRRTF 1235  
 QY 1019 TSGIRTSDDS-----SEEDPYM-----NDTVPTSPSADSTVLLAPVODSGSLN 1064  
 Db 1236 VNGTRQOPSWMLQATKSKRPIMLPVTFMIDGTTKTLTDSATTAELCALSDKSLDQ 1295  
 QY 1065 -----SSGSESTYCMFONAGDLPSPDGDYDQDDYDGAITSSSVTFSN 1110  
 Db 1296 RFGFSLYIALFDKYSSLSGSGN-----DHVMDAVSQCEQYAKEQQAQRNA---PWRLFPRK 1348  
 QY 1111 SYGSGSPDYRCVGT---YN-----SSGAYRFSSEB-AQSSFDESEEDPDRPPTDDE 1160  
 Db 1349 EITFPHNDPAEDVATNLITQOIVRGVKEGYEYCDREDLRLASQOYVYDGSSEILVERL 1408  
 QY 1161 L-----SYRBD-SYVSCVTLPPYFHSPLY-----MKGGLMNSMKRRMVCVL 1198  
 Db 1409 LSIIPSYIDPRELSSAKTYERMAQFIAMAHKKGYTQKVDPOKVEEVDFARYMPPL 1468  
 QY 1199 KDETFLMPSKQKALQGMHLHKKGGSSSTLSRNNKKRMFVLROS---KLMYFENDSEB 1254  
 Db 1469 FSRFYEAFFKSGSPLEK-----NDVVAANVMGVFVDEQEOVLLELSPETIAVS 1519  
 QY 1255 KJLG-----TVEVTKAKIIDNTTKENGIDIMADRTFLHESBEDSOMSVLSQVH 1308  
 Db 1520 SSKGGLQAOSEFTLATITKADEFTTSNNMEDIRDLVTF---DEGLRKSKEFVAL--- 1572  
 QY 1309 ASTDOEIOEWHDQANPONAVGTLVGLIDSVCASDSDPRPNSFVITIANRVLHCHADTP 1368  
 Db 1573 -----QDNP-----SPAADS----- 1583  
 QY 1369 EEMHMITTLQNSKQDTRV---EGGEFIVRGWLHKEVKNSPMSSKLKQRM----- 1417  
 Db 1584 -----TFLSFLKGLIIVLDQDTGEQVMTSGWAGH-----TMDRTKQRPDPADCV 1628  
 QY 1418 FVLTNHSLOYKSSERNAKLGTLVNSLCSVVRPEBKIFKETYNNVNTVYGSKHCYRL 1477  
 Db 1629 YVLP-----TVVRPHDV---VAFVTMTDQKQESLRT- 1658  
 QY 1478 TKLLNEATRWSSVIONVDTKABIDPTQO-----LIQIKENCLNSDVVEQIYKKNPIL 1532  
 Db 1659 -----SHVPALETBERVKNPYTLEEFSDYDFRPPKATILSRWITKARGDKLM 1707  
 QY 1533 RYTHAPLHSPLELPYGDINLNLKDKGYTTLQDEAIKIFNSI-----QOLBS 1580  
 Db 1708 CCTREPIKQALLK-----KCGHEHLSQEAQMAFIAVMKGYDPSKRTSVNE 1756  
 QY 1581 MSPRIFIIQILOTGHDARLDELCOULKO---TKVHPRGSVGLYXWQIITLCSCTF 1638  
 Db 1757 LTD-OIFEGALKA---BPKOEIYQILIKOLTEHNIKSEBK---WELLMLCVGLF 1806  
 QY 1639 LPSRGLIKLKFHLKRIREQPQTEMEKVALFTYESLKK---FKCREPVSREIBALH 1695  
 Db 1807 PRSNVLLPHVQRLQS-KKHP-----LALDMQQLQALRNGSKRPYHVLVEVRIQH 1859  
 QY 1696 R-OEMTSTYVCHGSGCKITINSHTAGEVEKULINGLAMEDSRNMFALF----- 1744  
 Db 1860 KTIQIFHKVYFPDPTBEAFESSTAKPCLNISRLTKTPEG-FSLFKISDKIVISV 1918  
 QY 1745 -----EYNGHDXAIESKTVADVLAKEKLAISBVODLPWKRYF---KLYCFIDJON 1795  
 Db 1919 PBGDPLDFVRHLTDMIKKARPAKDI-----VPSLTYQVFMKLM---TTT 1963

QY 1796 VP-KOS-VEFAFMFOAHNAVIGHHPAPEENLOVALAI--RLQYLQGDYTLHAIIPLE 1851  
 Db 1964 VEGKDSFADSIPIHYQELPKYLRGYKCSREEVFOLGALITYKVE-EDDKSHFPSIPKM- 2021  
 QY 1852 EYVSLQRLKARISOSTKTTTPCERLEKRTSLBEGTLRSFRFGSVNQKVEEQMLDMW 1911  
 Db 2022 -----LKEMTPO-----DLIRQLSPD-----W 2039  
 QY 1912 IKEEVSARASIIDKRRKFOGNORQAMAKYMALIKEMPGYSTLPDV-ECKEGGPOBL 1970  
 Db 2040 -----KRSIVAFENRAGSKREBAKMLFKIIFKMTFSGAFPEVQTEPRHPEL 2091  
 QY 1971 WGVASADAVSYKRGREPLEVQYEHILSPAPLANTYKIV-----VDERLEFETSV 2025  
 Db 2092 LIAINKVSVLIDPKNKDILITTPYTKISMSS--GNTYFHTIGNLVQSGSKLCTSLG 2149  
 QY 2026 DVVAKLMKAYISMVYKKRSTRSSSQSS 2056  
 Db 2150 YKMDLLTSTYSIOMLT--TMSKORNSRGS 2177

RESULT 8  
 ID P91443 PRELIMINARY; PRT; 2098 AA.  
 AC P91443;  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE T10H10.1 protein.  
 GN T10H10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berke M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Frevello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat V., Wohlman P.,  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Nelson J., Langston Y.,  
 RT "The sequence of C. elegans cosmid T10H10."  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.,  
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U00848; AAB37988.1; --  
 DR HSSP; P08799; 1MDN  
 DR InterPro; IPR000299; Band\_4.1.  
 DR InterPro; IPR00048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR InterPro; IPR000857; MYTH4.  
 DR InterPro; IPR000159; RA\_domain.  
 DR Pfam; PF00612; IQ\_4.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00784; MYTH4; 2.  
 DR PRINTS; PR00193; MYOSINHEAVY.



Oy	1669	IEALILRH-OEMTSTVYCHGGGSCKITNSTHTTGEVVEKILIRGLAMDSNMFPALFEYN	1747
Dd	1767	EVEALIOHKTTOIFKHFVPPDNTDEALIEVDSATRADPCHKIGYRLGLSS--DGSLSF---	1822
Oy	1748	GHWKATIESRTVADVLAKEFKLATSEVGLPMKFEYKLYCFLDTONVPKDSVEFAMF	1807
Dd	1823	-----VKIKDKVLA-----	1841
Oy	1808	EQANEAVYHGHHPAEF-----NLQVLAALRLOYLQSDYTLHAIPLLEEVYSIQRLKA	1861
Dd	1842	VRSLSDWHTHAKATQKADATMIPINQVYPMKLMWY---NFAVAGADPADIIIFYHQ---	1894
Oy	1862	RISOSTKFTFCERLEKERTKATSFLEGTLRSRFSRTGSVVRQKVEBEOMLDMWIKEE-----	1915
Dd	1895	---ESQKLLGYNHKTTKDVIETLALIRSTMKGOKAPLAQIPLDLDEIIPKSLKMY	1951
Oy	1916	VSSARASIIIDKMKRQFQGNNOBAMAKVALLKEMPGYSTLFDV-ECKEGGFPOLMLGV	1974
Dd	1952	ASEMRKTISSNAVRIEHLKSDQAKIEFLINYICRWPTFESAFPPASQVSDMLPRLILLAI	2011
Oy	1975	SADAVSYKRGEGRPLEVFQYEHILSFAPLANTY-KIVY-----DERELLETSEV	2026
Dd	2012	NOTGVNITHLDTKMLLVQPPFVNICMNTS--GNTYFMNTVGNMLKGNKGKILLDDTVGX	2069
Oy	2027	DVAKLMKAVYSIMVYKKRY---STRSAS	2051
Dd	2070	KMDLLTSYISLLISNONNHPSKTREVA	2097
RESULT 9			
ID	O9VLZ3	PRELIMINARY;	PRT; 2129 AA.
AC	O9VLZ3;		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	CG6976 protein.		
GN	M02881 OR CG6976.		
OS	Drosophila melanogaster (Fruit Fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephyrididae; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Gelinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoising R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays J.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dudlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabrielian A.E., Gary N.S., Galbart W.M., Glaeser K.,		
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,		
RA	Jarail M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Laeso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milblina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pauleb J.M.,		

Query Match	15.9%	Score 1705;	DB 5;	Length 2129;
Best Local Similarity	25.7%;	Pred. No. 4.1e-81;		
Matches 598;	Conservative 392;	Mismatches 828;	Indels 508;	Gaps 84;
3	LOGTRWMLRENGQHFSPSTVNSCAE-----GIVFRPTYGGVFT---- <td></td> <td></td> <td></td>			
13	IQGGYVWVKRPQ-----NTTSFVAVPFGARIVRTKQTQTLVCDNRKKQFWVAGDYLK 64			
53	AMHPTNEEGVDMASTELHGSGIMVNLFORYKKNQIMTYIGSILASVNPQIAGLYEP 112			
65	AMHTTSGEDVEDMTLDDLDQEYTLIRLQLQNRVYAKQLYTTGSMLVAINPQ-I LPTYN 123			
113	ATMEQYSRRHLGELPHIIPAIANECYCLMKRDNQCILKGSAGAKTSTKILKFLS 172			
124	REIOLYNNKSLAEPLPHIIPAIISDAFORLRLKENQCVCVIGSGEAGKTESTKILLYLA 183			
173	VISQGSLELSKEKTCSEVERALIESSPIMEAFGNAKTVVNNNSRFGFVQNLCOGNI 232			
184	AIS-----GHSWTEQDILFANFIMEAFGNAKTVVNDNSRFGKYIIRFTPGQAI 234			
233	QGGRIIVDCILSSQNRVVRQNGERNYHIFYALLAGLEBHEREREYL--STPENYHYINQS 290			
235	QGAIIQGYLL-EXSRIVFGSDENNYHIFCMLLAGLSTAEERAKLQEQSPQHYHLAG 293			
291	GCVEDKTSIQDESREYITANDVQMSKEBEYREVSRLLAGLILGNIETITA-----GGA 345			
294	GCFTLPGRGAKDPADIRAMKVLVSFKPEEWSILSLAAILHGNRFTATEVANLATA 353			
346	QVSKTKLAGSABLLGLDPQUTDALQORSNFLGSEIILPLVNOQA VDSRDSLAMLAYLA 405			
354	EIDTTPMLQRAVQQLGIPISALNALAQRTI PVNGEIVTTSLSKEAALIEGDAVKSLEYD 413			
406	CCFEWVVKIKINSIRIKGNE--FKSIGILDTFGFENFENHPECPNINYANEKLOEYFNKH 463			
414	GIFRIVRIRINETINKQVDPMNSIGVULDFGFENFDNNSEQUCINAYENLQGFYGH 473			
464	IFSLEOLEYSREGLWEDIDWIDNGECLDI- EKKGLALALINESSHPOATDSTLLEKL 522			
474	IFKNEQDEYQNEHINMGHIERQDNQQLDILGKPKMYLMSLIDBSKFPKGTDTQLLEKL 533			
523	HSQHANNHFYKPRVAVVNN- FGKVIYAGVQVQYVGRGLEKKRDTFRDDLNLALBSRDF 581			

```

Db      534 HVOHGHSIYVKGKTQTSIFGIRHAYGVMMNPLFLEBNRDSFGDGLRTTLQORSTNXX 593
Qy      582 IYDLFPHVBSRNNOPTLKCGSKHRRPTVSQPFKVDLSHLMATLSSNPFVCIKPNQ 641
Db      594 LVDFPH---EMPMDTAK-----KOPTLCYKFR-NSLDMIMRTLSQHFYFICIRPNEY 644
Qy      642 KMPDQDQAVNLNQLRYSGMLFTRIRKAGYAVRPFODIYKRYKYLMBRLALPE--DVR 699
Db      645 KKPNDKDELQVQLAYSGMMETARIRBAGYPIRHAIVRAVERIRLLVPVGLBQCDOR 704
Qy      700 GKSTLSLQYDASNSMQLGKTVFIRBSLEQLEKREHVSHAMVIRAHVGLPLAK 759
Db      705 KLAQICEVALPADSDROYGKTLFLRDEDDASLEQSQMLKSIIVTIQIRIRVLRFR 764
Qy      760 QYRKVLQCVITIQKNAFLRRRFLIKK-----AAIVQKQL----- 798
Db      765 YMKRYREAIITVQRYKRGLOKRRKYOVMMKGFHRLGACIAQOULTYKFTVRCRTIKLOA 824
Qy      799 --RGQIARRVYRQLAEKEQEKEKQOE-----EEEKKKREBERERER 840
Db      825 LBRGYLVKDFQKLEKRRKQNKKEELLKLAKEAEHLRLQOLKQKEREHQEOE 884
Qy      841 ERREBELRAQOEETKQOELEALQSKQKEAEITRELEK---OKEN----- 883
Db      885 KTLQESORLKAEMALMALMAVAVOOKRRTPVKQBPAPLPTLQANSLPPTTLIVA 944
Qy      884 ----KQVEILRL-----EKEIEDLOR-MIEQOELS-----LTEASL-- 915
Db      945 PLPTPASAATKRNITPESPRTIDVSSKOMVDVFRFLNDEBVGPLGPNKEXKSNMF 1004
Qy      916 -OKLQERRDOELRLLEEACRAQOEFLSINFEIDECVY-----NIERS----LSGG 963
Db      1005 EQALRRRROVPTKLISRPA-----LRKLNINSSODTILRLPSVNNIDTDSFSLKXA 1058
Qy      964 SESSSLASACEKEKNFNFSPQYREEVEDEGEADDAFKOSPNPSEHGHSQDRTSGI- 1022
Db      1059 ATYFGGATQAOHERKP-----LKKSILKHEPIDEMASKAIIW 1095
Qy      1023 ----RTSDSSEEDPYMNDTV-VPTSPASDTYLALPSVQDSGLNNS-SGSESTYCMQ 1076
Db      1096 LTTIRMGD-----LPDVSSPTLHFDNENLMS---DLASLINTSDSYKPLRFLVQ 1144
Qy      1077 NAGDLPSPGDYDYPDDYEDGALITSGSS---VTSNSNG-----SQMSPDY 1120
Db      1145 SQGRIFKPLASGKEAQEFYQHMLANVPTLHEKIHIIIGGIIKNSLRCPPEKEPERHL 1204
Qy      1121 KCSVGTNNSGAYRFSSEG-AQSSFEDSEEDPDSRPTDDELISYRDSYVSCVTLDPFHS 1179
Db      1205 R-----SFMKQGTALQATPRLQRLERTLVNGPRC--QPPSLFELHAIKGRHP 1250
Qy      1180 PLVMKGLNMSWRKRCYLKDETFMLFRSKOELKO-----GMLH 1219
Db      1251 -RLDILHMDQOQRRLQVDAAST-----AREVNVNLCQSMGLTDTFGGLVMSLNGKLM 1303
Qy      1220 KKGGS-----STLSRRYMKRW--FVLROSKLYVFENDSEBK-----LK- 1257
Db      1304 PLGAGEHVLDAISSECEQOLDAFMYLIRKEMFAWYDSDMNPXKATOLYKQILNGLKC 1363
Qy      1258 GIVEVNTAKEIIDNTT---KENGIDIMADRTFHLIAESPBD-----ASQMSVLS 1305
Db      1364 GEYCRSEBKDIAMVNCALACFEYEGPEILRLKSEITAFVPSDLAPGERAIEBMSRLIA 1423
Qy      1306 QYHASTDQEIQEHMDQANPQNAVGLDGLIDSVACSDSPDR---NSFVITIANRVL 1361
Db      1424 ATYKESYKVEENDLLLEAQKAK-----EDICLFAILSWPMHSHRLFVV----- 1470
Qy      1362 HCNADTPEEMHMTITLORSKQDTRVEGOFIYRWGLHKE-----VGNSPKXMSLKLKXR 1416
Db      1471 --RKEGRPKQSDMLMGINSAGFLIDETQVLAASCFBSVLKVNHESDDKLHMTFQYV 1528
Qy      1417 WPLTINS-----LDYKXSEBKALKLGTLVLSL--CSVVPDEKIPKETGYW 1463

```

```

Db      1529 NEVLQSSAODANEVINTMLDNLQRSSYGVALDPVEGDELDCLVINGDLIFEBAGVT 1588
Qy      1464 NNTVY--GRKHCYRLYTKLNEATRWSSVIGVNTDTKAPIDTPQOOLIDIKENCLN--- 1518
Db      1589 GAOUMAGNAQDCYR---GCYN--GOMGQFLAGNVKVLATLTPRSKLDQDILBGRQOEP 1643
Qy      1519 -----SDVEQIYK-----RNPILRYTHNPLHSPRLPLPYGINLNL 1556
Db      1644 KETPRANYSRRRQHNISQLAESHPRELDSDKAPLSKFSPEPLKAPIL----- 1691
Qy      1557 KDKGTYTLODEAIKIFNSIQ-----LEMSD-----PIP---IIQGLQCHDR 1599
Db      1692 -----KAVVXKPPLEQALVMMHNLTKMGDIARSNLVNTDLIFQPLQ--HPL- 1739
Qy      1600 PRDELQCOLIKQTNKVPHGSVGNLYSQILTSCFLPS---RGILYTKLKFH----- 1651
Db      1740 -LCDELYCOLMQLS--DNPSSESEKRGDILYLATGLVABSVLVKRELITILBRADYL 1796
Qy      1652 ----LKRIEOPRGTEMEKYA-LFTYESLKKTKC-----REPVSRLIEBALIHQEMTS 1701
Db      1797 ADACLKRLKRLSLAQOGRKAPHLIEGIIQ-RLCHYHKIYFPD-DVTEA----- 1845
Qy      1702 TYCHGGGSCKITINSHTTAGVEVEKLIRGLAMEDSRMFLPBYNGHYDKAIESRTVYA 1761
Db      1846 -----FEISHTREGAELIADIQRLBK-SPVQSYSLFKTGGRVVMPEEEFVP 1893
Qy      1762 DYIAK----FEKLAATSEVGDLPWKFPYKLYCFGLTDNVPKQSV--KFAFMPEQAHVAI 1815
Db      1894 DFTITQILYLRQORTIRISDQYQHFMRKML--NNRPGEDLNGDMIFSPOELAKYL 1951
Qy      1816 HGHNPAREBNLOVALRALQYLOQDYTLHAIPLREEVYLSQRLKARISQSTKTPPCR 1875
Db      1952 KGYVPIDEQASRLAIL-----YASADHVSILQRLP----- 1982
Qy      1876 LKRRTSFLBGLTRSFRTGSVVRQKVEEOMLDMKIEVSASASTIDKRRKQGANQ 1935
Db      1983 -----EVLTRLIPE---DLPLQTVABWKOQLIPKYNR--DHLE 2017
Qy      1936 EOMAKYMALIKEMPGYSTLPDV--EKEGSGPOELMGVSADAVSVYKRGSGRPLEVQ 1994
Db      2018 DHAKILFLQELSHFACFGSTFVVKQONDALPETHLIINISGTGHLDPTTKELISYE 2077
Qy      1995 YEHILSFGAPLANTYKI---VDERELLFETSEVVDVAKMKAYI 2036
Db      2078 YGQ-LGIWSSGKNHPIIRGNMIGASKILCSTTGQYKMDLLASVY 2122

RESULT 10
Q9NH54
ID Q9NH54; PRELIMINARY; PRT; 2121 AA.
AC Q9NH54;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Myosin VII-like protein.
GN MYO28B1 OR CG6976.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Anderson J.B., Yamashita R.A., Sellers J.R.
RT "Complete cDNA for an unconventional Myosin (Class VII) in Drosophila
melanogaster."
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233269; AAF34810.1; -.
DR HSSP; P08799; 1MD.
DR FlyBase; FBgn0040299; Myo28B1.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.

```



```

QY      1663  EMEKKA -LEFTESLAKTGC-----REFPSPDEIEALHNRQEMSTVYCHGGSGOKITN 1716
Db      1804  ORKKAETHLIEVGEIOQ -RCLNIYHITYPPD -DTEA-----FEIE 1841
QY      1717  SHTTAGEVVEKILRIGLAMEDSRNMALFEYNGHVDKAIESTRTVADVAK---FEKLA 1772
Db      1842  SHTRAEELIADIAGLELK -SPVGSITLKTGDRAVYAMPREEFVDFITOLLIMRQORT 1900
QY      1773  TSEVGDLFPWKYFKLYCFPLDTDNVEXDSV -EFAFMFOAHNAVIGHNPAEENLOYLA 1830
Db      1901  IRSISDGOYLHPMKMLN--NNHGEIDANGMITSYPOELHKYLGXYPIDCEQASRLA 1958
QY      1831  ALRLQLOQDVTTLHAALPPELEVYLSQNLKARISGSTFTTPCERLEKRRISFLEGTUR 1890
Db      1959  IL-----VYSADHDVSLORLP-----1974
QY      1891  SPRTGSVVRQKVEEQMLDMWIKKEVSSARASIIDMKRFGQOMNOEQAMAKYMLIKEM 1950
Db      1975  -----EVLTRLRPE---DLRLQYVAEMRQOILPKYR -DHLYEDHAKILFLOELSHFA 2024
QY      1951  GYGSITLPDV -ECKEGSPQOELMWGVSAVSVYKRGEPLEVPQYEHILSPGAPLANTY 2009
Db      2025  CFGSFTFVVKQONDALPETHLIALINSTGFHMLDPTTKELIRSVYSQ -LGIMWSGKNHF 2083
QY      2010  KI-----VYDERELLEFTSEBVVDVAKLMAYI 2036
Db      2084  HIRFGNMIGASRLCSTTQGYGMDDLLASV 2114

```

RESULT	11		
099M26		PRELIMINARY;	PRT, 2113 AA.
ID	099M26		
AC	099M26;		
DT	01-JUN-2001	(TREMBLrel. 17, Created)	
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Myosin-VIib.		
CN	MYO7B.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21295050; PubMed=11401444;		
RA	Chen Z.-Y., Haason T., Zhang D.-S., Schender B.J., Dertler B.H.,		
RA	Moosker M.S., Corey D.P.;		
RT	"Myosin VIb, a novel unconventional myosin, is a constituent of		
RT	microvilli in transporting epithelia.";		
RL	Genomics 72:285-296(2001).		
CC	-1 - SIMILARITY: CONTAINS 1 SH3 DOMAIN.		
DR	EMBL; AF242411; AAK28339.1; --		
DR	HSSP; P08799; 1IND.		
DR	MGD; MGI:107709; Myo7b.		
DR	InterPro; IPR002106; AAcRNA_11aase11.		
DR	InterPro; IPR000299; Band 4.1.		
DR	InterPro; IPR000871; Beta_lactamase_A.		
DR	InterPro; IPR000048; IQ region.		
DR	InterPro; IPR001609; myosin_head.		
DR	InterPro; IPR000857; MYTH4.		
DR	InterPro; IPR001452; SH3.		
DR	InterPro; IPR005058; S1gPase.		
DR	Pfam; PF00612; IQ; 5.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	Pfam; PF00784; MYTH4; 2.		
DR	Pfam; PF00018; SH3; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	ProDom; PD000355; myosin_head; 1.		
DR	SMART; SM00295; B41; 2.		
DR	SMART; SM00015; IQ; 4.		
DR	SMART; SM00242; MYSC; 1.		
DR	SMART; SM00139; MYTH4; 2.		
DR	SMART; SM00136; SH3; 1.		

Query Match	Similarity	Score	DB	Length	
Best Local	57.7%	1635	DB 11	2113	
Matches	577	Conservative	385	Mismatches	830
				Indels	542
				Gaps	82

Query Match	15.24;	Score 1635;	DB 11;	Length 2113;
Best Local Similarity	24.78;	Fred. No. 2e-77;		
Matches 577;	Conservative 385;	Mismatches 830;	Indels 542;	Gaps 82
QY	5	GTRVLRNGQHPSTVNSCAEGIVFRTDYGQVETTYKQSTITTH-----QKVTAMPT	57	
DB	7	GDHWLDPSP-----SKTGVAIGGI VETKTKGKTLIEDBEKHHVHADLSTLRPMHPN	62	
QY	58	NEEGVDMASTLTELHGGSIMYNLFQRYKRNQIMWTYIGSILASVNDYQPIAGLYEPATBEQ	117	
DB	63	SAQGVDDMIRLIGDINEAGVGHNLIRYQOHKITYTGSILVAVNPFQMLP-LYTLQEQVI	121	
QY	118	YSRRLHGLPHPIPAIANECCYRCMLKRNDCILKGGSGKSTGLKLKLSVISQ	177	
DB	122	YYSRMGELPHPIPAIANSCYFNMKNNRDCCIISSGSGKLETTLKLILOPLATVSGQ	181	
QY	178	SLSLKTKTSCEVERAILESPPIMEAFGNATVYNNSSRFGKFPVLNLCQKNGIGSRI	237	
DB	182	H-----SWIEQVLEANPILFAFNATKIRIDNSSRCKYDIHFNSSG-VLAGAS	231	
QY	238	VDCLISGNRVVRQNGERNYHIFYALLAGLEHEEREEFYLTSPENHYLNQSGVEDKT	297	
DB	232	IEHFLFKSRVCRQAPFEEKNYHIFCYMLGMGSPBEKQMLSGMPSEHYHLTWGSCTSSEG	291	
QY	298	ISDQSFREVIYTDVMDQFSKEEYREVSRLLAGLHIGINIPITL-----GGAQVSFKTA	352	
DB	292	LSDADVDYHVSAMKLIQFSDSENNWDISKLLAAILHLNCGMAVAFENLSSDMETPETA	351	
QY	353	LGRSAELLGLPTQLTDLATQSRMFLRGEELTPLVNQAADVSRDSLAMALVACCFEWY	412	
DB	352	FPLMKLELVQHQALRDLCLIKHTTIPVLGEFVSRPNINQADRRAFAFKGIYGRLEFQWIV	411	
QY	413	KKINSRI-----KQNEDFKSGIGLIDIFGFENFEVNHFEQFNINYANEKLOEYFNKHIFS	466	
DB	412	KKINAIPTPOADQPNVRAIIGLLDIGFENFQNSSEQLCINPANEHLQOFPVGHVPT	471	
QY	467	LEOLEYSBGLWELIDIMIDNGECDLIE-KKLGLLALINBSHPPQATDSTLLEKHSQ	525	
DB	472	MEOEYLSBENTWNYIHTDNOPLIDMLAKPMSIISLDESESPQCGDVTMLQKLSNI	531	
QY	526	HANNHFVYKPR-VAVNNFGVGVHAGEVOGYDVRGLEKORDFPRDOLLNLRSEPDFIYD	584	
DB	532	HANNKSFSPSISIDHTRFGIAHFAQDVYYQAGFLFKRDRDLSTLILIHSSKQKFLKE	591	
QY	585	LFEHVSNN-----NODTLKGSKRRP-TVSSQPKVDSLHSLMATTSSNPF	631	
DB	592	IFNVDSQTLGHGHTICQKAGSGQLFKSSDSIKRPVTLASQPK-OSLDQMLKILTNQGY	650	
QY	632	FVRCIKPMQKPPQDPDAVVLNQLRYSGLMETVIRKAGVAVRRPDPFRYRYKVTLM--	689	
DB	651	FVRCIKPMYKPPFLFDELELCIQOIRYSGMVEYAIRKSGPIRATYFQFSQFRFLPS	710	
QY	690	-RLNLALPDVGVGKCTSLLOLVASNSBWLQGTKYFLRESLEQKLEKREBEVSHAM--	746	
DB	711	PERMQFQKPKRQMTIHIDLCLGTDKMKVGTKIFLQHDQDVTVEIRSQALDGAIRI	770	
QY	747	--VIRAH-----VLAGF-----LARKQYR	762	
DB	771	QVLRGHKYRKEFLQRRAAVTLQAGRWGYSQKKNFKLILVGFERLQAIARSHLLMRQPK	830	
QY	763	KVLVCVYLIQKNYRFLRRRFLHKLKAAIYFQKLRQOQIARVYRQ-----LMAK	814	
DB	831	AMRORIVOLQARCRGYLVROQVQAKRRVAVIIOAHARQVAVRKSTWQCKSTGPVILA--	888	
QY	815	REOEKKQOEEBKKGREEEERERERE-----AELPAQOEEETRKOQOELALQKS	867	

```

Db      889  --KEPAAQVAVHERKRSIYDTVDTAMVEKVFGLPAMIGCGECPAPTRFEDLEV--KT 944
Qy      868  OKAEELTRELKQKEMKQVEILRLKEKIEDLORMKEQOELSTE-----A 913
Db      945  OKKHE-----VDDDTVPMAMPEEEDVSLAETFPKFAVTVQKSA 985
Qy      914  SLOKORRDEQLRLKEEACRAA--QEFLESINF--DEIDECVRNTERLSGSGSEFSSEL 970
Db      986  SHTHIQKPLKVPPLLYHENNDTHSALDWIILIRFNGDLPBPVYVGRNSLTGSSVWR---- 1042
Qy      971  ABAACEKKNFNFSQRYPEEVEDEGEADDAFKDGSNPNSEKHSQORTSGITSDSSE 1030
Db      1043  -----QIHDKLGKDS--VTQHRSSQVLSQMLNFGGEARK 1074
Qy      1031  EDPYMDTVVPTSPADSTV---LLAPVOD-----SGLSNLSGSESTYGMQPN 1077
Db      1075  FDDPISDRPMNLKRNHFIVGAIKMRGLDEIYCOICQKLSNNTSSRPMWILLSLC 1134
Qy      1078  AGDLPPDPDGDYDODDYEDGATTSQS-----SVFSNSYCSQ--W----- 1116
Db      1135  LGGPPSERFMKTLNPIGSPSYGPFCAERLQRTFANGVRABPPTWLEQAVSKKH 1194
Qy      1117  -----SPDYRCVGTYNSSGAVRFSSEGAQS-----SEDESEDPDSRFPDDELST 1163
Db      1195  PIQVILATGRSLTISVDASSTSRICQHVAKQGLRDLNLFSLQVAVVDKFMSLGSGCDH 1254
Qy      1164  RRSVYSVCTLPRFHSFLWKGLMNSWKRMKCVLDEFTL--WRSKQELMKGMHLKXG 1222
Db      1255  LMDAVACQEQALABERG---ESQROAPWR---IYFKFEFTPMHDSQEDVSTELIYHO- 1306
Qy      1223  GGSSTSRNMKRWFLRQSKLMYFENDSEELKGVLEV--RTAKEIIDNTKENGIDI 1280
Db      1307  -----VLRGVMSGEVNFKEBEL---VELLAHCYVQIGATYKSNAYQE 1347
Qy      1281  IMAD-----RTFHLIASEPDAQOWFVLSOVNA-----STDQEIQENHDAQNP 1325
Db      1348  LRPSCVPSKLYRT-----KSPE---KMASLVTAHAKAOTQSKATPLAREQTEA-- 1397
Qy      1326  QNNAVGTLDVGLISVCASDSDPRPNPSPVITANRVLHCNADTPEBNNHTTLQRSKGP 1385
Db      1398  -RLMPLLSRLREVTTLSCGPRLPKQVLAIN-----WKGMFLDOKER 1441
Qy      1386  RVEGOEFT--VRGWL--HKEVKNSPKMSLKLKRWFLVTHNSLDYKXSEKNAKLGTLV 1443
Db      1442  TLGLSFAEVMGVVARNRDAQKKLLATLQEEYEFVSPSV---AIAEWNALFLGGLKE 1498
Qy      1444  NSLCSVVPPEKTFKETGVNNTVYG-RKHCYRLYTK--L-----NEATRWSSV 1490
Db      1499  RSVFAMALQDRATDD-----ITLPPKKGDLILTKKQGLLASEMWALQONDRGTGTL 1553
Qy      1491  IQNVTDTKAP--IDPTQQL-----ODIK--ENCLNSDVVEQIY----- 1526
Db      1554  VPTACLYTTPSVTKPSTQLSLSLAMSPBEKRLAAGEVRLPELDQLTSPPTLEEFST 1613
Qy      1527  -----KRNPIRLYTHNPLHSPLLRPLUGDINILNKLKQGYTTLOD 1566
Db      1614  QPRAPEKETISRGAAMPASRSHLWASPEPLRQL-----LKSVDK--AKLND 1662
Qy      1567  BAIRIF-----NSLOQLBSM-----SDPIPIIGLITQGHDLRLPRLDE 1605
Db      1663  AACQIFALILKYTGDPYPSRQSWHSLLELDMFSLALDP-----ALQDEL 1707
Qy      1606  YCOLIKOT--NKYPHGSVGNLYSNQLITCLSTCLPBRGILKYLKHRLIREQPGRE 1663
Db      1708  YCOLIKQLTNNSI---RFESEERAKQMLMLCTGLPFGKTLPH-----AQKIDSR 1755
Qy      1664  MEK-VALTYESLK---KTKCRFVPSRDEIBAL-----IHRQEMTSYVYCGGSSCKI 1713
Db      1756  KKRPLALDCSRRLHRLVLRVGRKQPHNDVVKAAEQNVSKLHNE-----VYLPNDISKM 1810
Qy      1714  TINSHTAGEVVKLIRG---LAMEDSRMFLFEYNHVDKAI-----ESRTVY 1760

```

```

Db      1811  EVGSSSRVRDLCEGI--GTRQLASWDGCSLEFIK-----IDKVISLEKGEFFEDSLRQV 1863
Qy      1764  ADVLAKFEKALATSEVGLPMKVFYKLYCFLDTONVPMDSVEFAFMQAEAVIHCHRP 1820
Db      1864  SDWVKKNRPQKEGASV--TLPIYVFEMKMLNIV--TPKQDN-----ADTIAHYQE 1912
Qy      1821  ABEENIQLVLAIRLOYLGQDYTLHAAPLPEBEVYSLORIKARISOSTYTFPCRLEKR 1880
Db      1913  LP-----KYLKG-----FHKC-----GRE 1926
Qy      1881  TSFLEGTLRSPRTSVVRQKVEEQMLDMWIKREVSSARASIIDKMRF-----QG 1932
Db      1927  DAHLUGGLICKIQFSDSSQLASVSKVLKEIVPQNLTLMS--BEWKSILLBEDCKNR 1984
Qy      1933  MNQEQAMAKYVALIKEMPGYSTLPDV--EKEGGEFPOEIMLVGSADAVSVYKRGGRPLE 1991
Db      1985  KTVABKAYEFLKMYRMVPIFGSAFVEVQTSRPSFDILLAINHGLVLIHPKTELIN 2044
Qy      1992  VFQYHILISFAPLANTYKIV---DERELIFTSEVVDYAKLMKAYISMIV 2040
Db      2045  TYPFTKISSWS--GNTYFHMALGSLGQGRLLCETSLGYKMDLLTSYVQQL 2096

RESULT 12
Q9XE14
AC Q9XE14 PRELIMINARY; PRT; 1529 AA.
ID Q9XE14;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Unconventional myosin heavy chain.
GN MYO1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu L., Pearecra T.C.;
RT "Myosin gene family of maize.";
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF104924; AAD17931.2; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD003176; DIL; 1.
DR SMART; SM00015; IQ; 5.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ
SEQUENCE 1529 AA; 173455 MW; 22BE5947DC3EB3D4 CRC64;

Query Match 14.1%; Score 1512.5; DB 10; Length 1529;
Best Local Similarity 34.6%; Pred. No. 3.7e-71;
Matches 387; Conservative 187; Mismatches 403; Indels 143; Gaps 26;

Qy 5 GTRVNLRE--NGCHPSTVNSCAEGLVVRRTDYGQVFTYKQSTITTHQKTAHNPTEECVD 63
Db 10 GSHVWEDPNLAMDGEVVISIKNNEVHVQTSNGKVTYDRSKVFPKDEA--PFG--GVD 65

Qy 64 DNASITLHGSGIMYNLFORRYKRNQIMWTYIGSILASVNPYOPDIAGIYEPATMEYSRRL 123
Db 66 DMTRLSYLHEPQVQLNLAIRYELNELYTTTGSILAVNPFQQLPHLYDTHMMEQYKGAADF 125

Qy 124 GELPHIPIAINECYRCIMKRHDNOCILIKGSGAGKTESTYLILKFLSVISQOGLSL 183

```



```

126 GELSPHDAIAVAVYAMAMNEGKSNLSVSGESGACKTEITTKMLKMYLHLGGRS---CV 182
184 KEKTSQVEEALLESSPIMEAFGNKTVYNNNSRPFKFCVCLNLCOMGNIOGGRIYDVCIS 243
183 EGRT--VEQVLESNVLHFGNAKTVRNNSRPFKEVLEQPDKGRISGAIRTYL-- 239
244 SONRVKONPGERNYHIFYALLAGLHEEEREFEYLTSPENHYLNSGCEVEDKTSIDQS 303
240 ERSRVQINSPEKNYHCFPLCAA--PEYTORKYSDDPSFHYLNGSSCIEVDGINDAE 298
304 FREVITAMVWQPSKEVEVSLGIIHLNIEETITGGAOVSKTA-----LGR 355
299 YATPRAMDIVGNEEEOGIFPVVAVALHGINF--AKGEIDSVIKDKSRFLMI 356
356 SAEELGDPDTOLDALTORSMPLRGEIITPLVNOQAVNSRDLAMALVACCEPWYIKI 415
357 AAEELKCCONLEKALITVYIYVPEVITRTLDPAVASRDLAKIITSRLFDWIVEKI 416
416 NSRIKCNEDFKS-IGLIDIFGEPENFEVNHPEOFNINYANEKLOEYFNKIIFSLQLEYR 474
417 NVSIGDPSNKOLIGVLDIYGFSPVNSPEQICINYTEKIQOHQNVHFKNQGEYLR 476
475 EGLWEDIMDNGECLDLIEKKGLLALINESHFPQATDSTLEKLSQHANNHYVYK 534
477 ERIINSYIEFVDNODVLDIEKKGGLIALDEACMFPRSTHETFAOKLYTFEKNRRFPAK 536
535 PRVAVNPFVKGHAGEVOYDVRGILEKNDFPDLDLNLRESRPFITDLPHEVSRNN 594
537 PKLSRTDFVAVHAGSVTTQADYFLDKNDYVVAEHODLNMASSCLFVAGLEPDL- 592
595 OPTLKCSGRAPTEVSSQFVDSLSHLMATLSSNPFVYCIKPNQKAPDQCAVNL 654
593 QETAK---SSKSSISGRFL-QLOSLETLSSTEHYIRCVKPNLRLPAIFENNVIQ 648
655 QLRYSGLTETRIKRGYAVRPFOPYKRYKYLNNLALPEDVRK-----CISLDQ 708
649 QLRGCGVLEAIRISCAGYPRKTFYEFVRFVL-----APEVLEGSNDKLAQKILEK 703
709 YDASNEWOLGKTQVLRSLBLEKREBEESHAMVYRANVGLFARQY----- 761
704 MGLN--YQIGKTKVLRAGQAMDLARREVEVGARARITIQRIITYIARQPAELKRA 761
762 -----RKVLYCV-----YIQNYPAFLRRRFLKKAATVFOKLRGQIA 803
762 TQLQSFVGTGLARKLYECMRQBAAYKIQKNMRHHRFESYVLOQAATITLQGLPAMA 821
804 RR-----VYRQLLAEKREGEKKOE 824
822 RKEFRPKETKAVHIQAOWRHRHDYSHYKNIQGAALTYQCAWRQRLARRELKLMAR 881
825 E-----BEKKREBEEREREREREARLRAQOEEFTKQOELAQSOKEAELTRE---- 876
882 ETGALGKADKLEKRAVELTWIRGLERKRLTDLSEAKADQIADLQSTLDMQVSESKA 941
877 ---LEKQKENVQEYLRLKE-----IEDLQ-----MQEOEELSTEASLOKQERR 922
942 MVYKERRAARKAIEEAPVYKEFVVEDEKINSITTEVEQKALLITERQATEAKHE 1001
923 DQELRLIEBACPAQEPLESINFDIDECVRNIERSLSCGSSSELSAECERPNFN 982
1002 HAESELNHELKIKFESAEEKI--EQLQETVHRLBEKATVSESNNKYLQQAVALISPTSK 1059
983 FSGPYRE-----EVDGFEADDAFADS-----NPSF 1011
1060 SLAAVFPSPQOLKTPENGALNGEVKSSPDITPLNPKE 1099

```

RESULT 13  
Q9VZ48 PRELIMINARY; PRT; 2424 AA.  
Q9VZ48, 01-MAY-2000 (Tremblrel. 13, Created)

```

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG2174 protein.
GN CG2174.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Burton R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davernport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov J.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaimel M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mierlov G., Mielshina N.V., Mobarry C., Morris J., Moshirif A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splitter B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-P., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
DR EMBL: AE003484; AAF47980.1; -.
DR HSSP: P08799; IAMD.
DR Flybase: FBgn0030252; CG2174.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR000857; MYTH4.
DR Pfam: PF00612; IQ_3.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00784; MYTH4; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ_3.
DR SMART: SM00242; MYSC; 1.
DR SMART: SM00139; MYTH4; 1.
DR SMART: 2424 AA; 267616 MW; 8AD62MA339PAA5D0 CRC64;
SQ

```

Query Match 13.9%; Score 1495; DB 5; Length 2424;  
Best local similarity 40.3%; Pred. No. 5.7e-70;  
Matches 347; Conservative 146; Mismatches 294; Indels 74; Gaps 20;  
QY 48 HOKVTAMP-----TNEGVDMASLTIELHGSIWYVLFQRYKRNQIWTYIGSL 97

```

Db 133 HPOTFALQPEBGSLRARQDLSSGVEDMTLLDLHNSALMLNRLRYKGLIYTFAGSIL 192
QY 98 ASVNPYOPAGIAYEPATMEQYSRRHGLPPIHFAIANECYRCMKHNDQCILIGESG 157
Db 193 IANPYMPFADAYGLVAKQYAGRPGLSLPPHIFALIAAAHAAH---PSQVIVISGSG 249
QY 158 AGTSTKTLILKFLSVISQSLSLSEKTSCEVERALIESSPIMEAFGNAKTYNNSSR 217
Db 250 SGTSTKTLVMOYLAAVPPGGSAS-----AVITEQLLEAAILLEAFGNARTARNDNSSR 304
QY 218 EGFVQVQINQKQNIQGRIVDCILSQNRVVRKQNPERRNHYFYALLAGLEHEBEFF 277
Db 305 FGRTLEYVF-KSGAIVAKITQYLL-EKSRIVYQAPERNYHVFYELGLSESTERSKYG 362
QY 278 LSTPENHYALNQ--SGCVEDKITSDOSEFREVITAMVMOFSKBEVREYRLLAGILHLG 335
Db 363 LLEADKFFYNQCATDASGRV--DNESLQ---GANOVLGVSCEGRGIVRVLLAAVHLHG 417
QY 336 NIEF---ITAG--GAQVSFKTAGRSALLGLDPTQLTALTQSRMFLRGEELTPPLVY 389
Db 418 NVYFHRQLRHQGEVGVSDAEIKMAAHLHLISADLHRALTSRTTEARAEHLHTPLGI 477
QY 390 QOAVDSDSLAMALYACCFEYVYKIKINSRIK--GNEPFKISGLIDTFGEFNEFVNHPEQ 447
Db 478 DQALDADAPAKALYAGLFNWLRSINSIYQKGTDAHRISILDFGFDLAENSPFOL 537
QY 448 NINVAENKLOEYFNKHFISLEQLEYSREGLVMDIDWIDGECGLDIEKK-LGLALLINE 506
Db 538 CIYVAENLQLYRNKHFVLEQAEYARELERLETPAMDNLPIYHILAKKPVGICHLDD 597
QY 507 ESHPEQATDSTLLEKHSOHANNHFYVPRVAVNNGVKNYAGEVOYVGLIEKORDTE 566
Db 598 ESNPRATDLSFLEKCHYNALSELVAPRIGAEFGVTHYAGQVWCYVDGFLDKRDAL 657
QY 567 RDOLLNLRSRDEFTYDLFEHV-SSKNMODTLKCS-----KRRPVSQOFKDSL 618
Db 658 RGVLELILASSRPLVGLLTKLQAOQADACKTLPKSGNGFVTMKPTTPVAARF-ADSL 716
QY 619 HSLMATLSSNPFVRCIKENKQMPQDQAVLNLQRYSGMLETVRIKAGYVARRPF 678
Db 717 QQLQSGRGCHPWFVRCIKENKQKALRMPCVLOQLRGLMDITQIRQRYPRRLRF 776
QY 679 QDFYKRYKVLNRN-LALPEDVRGKCTSLQLYDANSSE---WQDKTYVFLRESLEQLE 734
Db 777 QHFVERYRHLPLSPLAGRTYRELCTRALBAMPRTGVEGPDYQLGATRVFLREALHRALE 836
QY 735 KRBEESHAAMVIRAHVIGFLARKQYKLYCVIIQKRYRAFLLRRPLHLKKAIVF 794
Db 837 SGRTERLRRAVSVQRHVRCMLVLRQLARRQAAATRLQARWQORAQOQRYERLRKALTA 896
QY 795 QKOLRGQIARRVYRQLAEKREOEKKEEKKREBEEREREREREREAELRAQOEE 854
Db 897 QRLMRGQARRRRQQLRSDHRRQEAR-----BAQRAREARA----- 935
QY 855 TRKOQELIALQKSKK--AEL 873
Db 936 --KOAVLERSQLSYLDIPAL 954

```

## RESULT 14

```

QY 09JUN02 PRELIMINARY; PRT; 2548 AA.
AC 09JUN02;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Myosin-IXa.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99339979; PubMed=10409426;
RA Gorman S.W., Halder N.B., Grieshammer U., Swiderski R.E., Kim E.,
RA Welch J.W., Seabry C., Leng S.C., Carmi R., Sheffield V.C., Dull D.M.;
RT "The cloning and developmental expression of unconventional myosin IXA
RT (Myo9A) a gene in the bardet-biedl syndrome (BBS4) region at
RT chromosome 15q22-q23."
RL Genomics 59:150-160(1999).
RL EMBL; AF117888; AAD49195.1; -.
DR HSSP; P08799; 1MDN.
DR InterPro; IPR002219; DAG-pe-bind.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000159; RA_domain.
DR InterPro; IPR000198; RhogAP.
DR Pfam; PF00130; DAG-pe-bind; 1.
DR Pfam; PF00612; IQ_5.
DR Pfam; PF00063; myosin_head; 2.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00620; RhogAP; 1.
DR PRINTS; PR00193; MYOSINHAIV.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00015; IQ_5.
DR SMART; SM00242; MISC; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00324; RhogAP; 1.
DR PROSITE; PS00479; DAG-pe_bind_DOM_1; UNKNOWN_1.
DR PROSITE; PS50081; DAG-pe_bind_DOM_2; 1.
DR PROSITE; PS50096; IQ_1.
SQ SEQUENCE 2548 AA; 292703 MW; B93B76C2A0E9A356 CRC64;

```

Query Match 13.9%; Score 1492; DB 4; Length 2548;

Best Local Similarity 24.3%; Pred. No. 8.8e-70;

Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;

```

QY 35 YGVFTYKOSTITTHQKVTN---MPTNEBGVDNMASTLEKSGSLMTNVLFOYKXNQWT 91
Db 118 YGSLQSWLRVTEREERMMERGFLLPOQKDFDCLSLDNLKNTLENNLDRFKHKKIYT 177
QY 92 YIGSLIASVNPQPIAGIAYEPATMEQYSRRHGLPPIHFAIANECYRCMKHNDQCIL 151
Db 178 YVGSILIVINPKF-FLPTNPKYVKNYDNDHQKPEPHIYAADVAYAHMLQKKNQCTV 236
QY 152 IKGESGAGTSTKTLILKFLSVISQSLSLSEKTSCEVERALIESSPIMEAFGNAKTY 211
Db 237 IKGESGSGKTQSTNFIHLTLALSGKF-----ASGVEQLILAGPLAEFGNAKTAH 289
QY 212 NNNSSRFKGFVQVLANICQKNIQGRIVDCILSQNRVVRKQNPERRNHYFYALLAGLE 271
Db 290 NNNSSRFKGFVQVLANICQKNIQGRIVDCILSQNRVVRKQNPERRNHYFYALLAGLE 348
QY 272 EREEFYLSSTPENHYALNQ-----SGCVEDK---TISDOE---SPREVITAMDV 313
Db 349 ERSAPHLKQPEEYHNLNQITTKKPLRQSDNDYTYDEBPDCFTVEGEDLRHDFRLQLANEM 408
QY 314 MQFSKEBEVREYRLLAGILHLGNI-----EFTTAGAQVSKTALGRSALLGLDPTQLTD 369
Db 409 VGFLEPKTRQIFSLLSAILHLGNIQYKKKTYRDSIDICNPRVLPVLSLEVEKEMLPFE 468
QY 370 ALTVKRTYVVGSKTLIPYKLAELAVTVRNSMASLSALFMDWIVFRINHALNLSKLEHNT 528
Db 469 ALTVKRTYVVGSKTLIPYKLAELAVTVRNSMASLSALFMDWIVFRINHALNLSKLEHNT 528
QY 427 ---STIGLIDTFGEFNEFVNHPEQFNIYANKEKLOEYFNKHFISLEQLEYSREGLVMDID 483
Db 529 KTLSTGVLDIFEFDEYENNSFQFCINPANNERLQHYFNQHFLEKQOEBYRTGIGMHNID 588
QY 484 WIDNGECGLDIEKK-LGLALLINESHPQATDSTLLEKHSOHANNHFYVPRVAVNPF 542
Db 589 YIDNQCINLISKKPTGLHLHLDSESNFQATNQGLDKFKQHNDNSYIEPRAWEPRAF 648
QY 543 GVKHTYAGEVOYVGRGLIEKRDTPFDLLNLRSRDFDIYVL----- 585

```

Db 649 IIKHVAQKVKYKQFREKQTHMPDIYALLRSSKNAPISGMIGIDPAVFRMALRAF 708  
 Qy 586 -----FEHVSNN-----NOTTLKCG----- 601  
 Db 709 PRAMVAFRAGKRNIRKTKGHDTAPCALIKSMDSFLOHPHQHSLIILQCKEKKXS 768  
 Qy 602 ----- 601  
 Db 769 ITRKNPRTPLSDLOGMNALNEKNQHDTPDIAMNGRTGIRSLSGTSLLDKXGIFANST 828  
 Qy 602 -----SKHRP 607  
 Db 829 SSKTLERAHGILTRKNPFKSPALPHLEVNSLKLTRLTLODRITKSLHLHKKKRP 888  
 Qy 608 TVSSQFQVDSLHSLMATLSSNPFVRCIKPMQKAPDQFQAVVLNQLRYSGLMETVRI 667  
 Db 889 SISAOQA-SLSGLMETLGOAEFYKICRSNNEKPLRPSDLVLRQRLRYGMLTETVQI 947  
 Qy 668 KKAQIAVRRPPODFYKRYVILMRNLALPEDVNGKCTSLLOLYDASNSEWOLQKTVFLRE 727  
 Db 948 ROSGYSKYSFODFVSHFVLLPRNIIIPS--KENIQDFPRKINLNPQNYQVQKTMVFLKE 1005  
 Qy 728 SLEQLKREKEEV-----SHAAVYIRAHVUGFLAKQYRK- 763  
 Db 1006 QSRQHLQDILLHOEVLRRIILLORFVLLCRQHFLHLROASVILIQFWRNYLNOQVDRDA 1065  
 Qy 764 -----VLYCVIIOKNVRAFLRRRFLHLKKAIVFOKOLRGQIARR----- 805  
 Db 1066 AVOKDAFWASAAALLOASWRAHLERQYLELRPAALIVQAKRQDYYRRRHAAICTQAR 1125  
 Qy 806 --VYRQLAEKREOEKKKQOEKKKREBEREREREREAELRAQOE-----ETRK 857  
 Db 1126 WKAYRE--SKRYOEQRKIIILLOSTCRGFRAQRFKALKEQGLRETKREVLNIXGYG 1182  
 Qy 858 QOELELOKQO-KEALTELEKQENKQOYEILREKEIEDQPKF-----QOELSLTE 912  
 Db 1183 SLEIQSDSEWEDCSFDRNRKAIIECKSVIESNRISRB--SSVDCIKESPNKQOEBAQOQ 1241  
 Qy 913 ASLOKLOE-----RPODELRLLEEACRAOEFLESINPDEIDECVRIE-----RELSCG 963  
 Db 1242 SGVD-LQEDVLRERPRSLIEDLHQKVGRAKR--ESRRRREZLOALFSLLELKNLSLGGI 1298  
 Qy 964 S-----EFSSLEAESACER--KPNFNFSQ-----PYPEE-----VDEGEADD 1001  
 Db 1299 SPSEDRRMSTELVPEGLQSPRGTPDESSESGSLLELSEYEQSGKLESYISDEG-----D 1353  
 Qy 1002 AKRDSNPSEBHGHSDDRTSGIRTSDDSR-----EDPYMDDTVYPTSPSADSTVLAPSPQ 1057  
 Db 1354 LQPPSPKISSSPKFDNRNALASNETSSAHLKQDTMKEMVVCSSBS-----ITCKPQLK 1409  
 Qy 1058 DSGSLHNSSGESTYCMPONAGDLPSPDGDYDODDYEDDAITSGSVTFPSNYSQMS 1117  
 Db 1410 DS---FISNLSLFFFIPOODPLKINSQUDTSGRKKLBNEDTAEALITLDINRETR-- 1464  
 Qy 1118 PYRCS-----VGTYNSSGAYRPSSEGAOSSPEDESDEDP:REDTDEL----- 1161  
 Db 1465 -RYHSGKQOIVPSLNTESNPVLKYLKLTKEKERQOQOQONKEMMEQIROQDTIL 1523  
 Qy 1162 -----SYRDSVYSCVTLPTFHSFLYMKGGLNMSMKRNCVLTQDEFLMFRSKQF-- 1211  
 Db 1524 EKERRAKTIEKRICECLVAP-----SSYOSKQOVERPSSILSLINTSNKE 1570  
 Qy 1212 -----ALKQGLHKKGGSSSTLSRNNKGMFWLROSKLWFEFNDSEKLGTV----- 1260  
 Db 1571 LNVLSLSLKDALLAQDSSSAHLPPKD-----RNVYVFEKKSFPQSSSTYKELS 1621  
 Qy 1261 -----EYRTAKEIIDN-----TTKEN-GIDI 1280  
 Db 1622 KTDRCMGTQJLVACKLNNRISKREHPRPQSYSHNSDDLSREGNARPIFTPRDNNSITL 1681  
 Qy 1281 IMADRTFLIABSP-----EDASQW--FSVLSQVHAATDOE:QEMHDEQANPONAVCTLDV 1334  
 Db 1682 VSKEA--LNSKNPQLHKEDEPAMKPVKLAGPGQRETSPRPSSV--DEQAKLHK--TMSQ 1734

Qy 1335 GLIDSV-----ASDPDRPNSFVITIANRVLHCNADTPRENNHMITTLQRSKGDTRV-- 1388  
 Db 1735 GEITLAVRQKASBDBIRPQ-----AKORFMAKQOGEKKTTRVAPT 1777  
 Qy 1389 -----GOEFT-----VRGMLHKEVONSPPKMSLSLKKRW 1417  
 Db 1778 TQSEVSPFAGTDVPAHQFDELAAYHPTPPLSPBLPGSCRKERKENKESP----- 1830  
 Qy 1418 FVLTHNSLDYKSSSKNALKGLTVLANSLSGVPPDEKIPRETGYMNTVYGRKHCVRLY 1477  
 Db 1831 -----KAKKRSVKISNVALDSM-----HMQ----- 1851  
 Qy 1478 TKLNEATRWSSVIONVTDTRAPIDTPQOOLIODI-KENCINSDVVEQIYKRN-PILRYT 1535  
 Db 1852 -----NDSVQ-----LIASVEDKS-MDEFLKKNVDLNDNSKDDTLVDVFPKALKPRQN 1903  
 Qy 1536 HHPLHPLPLPYGDININLKDQGYTTLQDBAIFNSLQOLESMSD-PIPIIGILOT 1594  
 Db 1904 IFSPYSALAMDQ-----KSIRYKDLVALFEQILEKTRLEORUSLIGSPYRV-----W 1953  
 Qy 1595 GHDLARLDELXCOLIKOTN-----KVP-----HPSGVNLYSQILT- 1632  
 Db 1954 VNTFVFLDE--YMEFKTSDCTATKVPTEKRRKRRKSTDLVEBHNGHIFRATQYSIPTY 2012  
 Qy 1633 CLSCTFLPSRGLIKYLKRIEQLREOPGTEMEKYLFTYESLTKT-XC-----R 1681  
 Db 2013 CEYCS-----SLWIMDRASVCLCKYKACHKKCKLKTATACKSKYDELSR 2059  
 Qy 1682 EF-----VPSRDE-----IEALHROEM-----TSTVYCHGGGSCITINSHTTAGEV 1726  
 Db 2060 QFVGLSRLTSEBDRTPVLVVEKLINVIEMHGLYEGIRYKSGSTWKI-----K 2107  
 Qy 1727 KLIRGLAME-DSRNMFALFEYNGHVDKAIESTVADVADLAPKELAAISEVGDLPWKRY- 1784  
 Db 2108 ELRQGLDTPDASVN--LDDYNIR-----VIASVFKOM-----LRDLNPPLMT 2147  
 Qy 1785 FELY 1788  
 Db 2148 FELY 2151

RESULT 15  
 QRYE8  
 ID QRYE8 PRELIMINARY; PRT, 1611 AA.  
 AC QRYE8:  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Putative myosin heavy chain.  
 GN AT2G33240.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCB1\_TaxId=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounseley S.D., Ketchum K.A., Lin X., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Kaul S., Mason T.M., Kertavage A.R., Adams M.D.,  
 RA Somerville C.R., Venter J.C.;  
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.

DR ENBL, ACC02334; AM14807.1; -.  
SQ SEQUENCE 1611 AA; 183741 MW; 98221A9494D60679 CRC64;

Query Match 13.9%; Score 1487.5; DB 10; Length 1611;  
Best Local Similarity 32.5%; Pred. No. 8; 1e-70;  
Matches 401; Conservative 194; Mismatches 378; Indels 259; Gaps 35;

QY 5 GTRVWLR-----NQHPSTVNSCAEGLVFR-TDYGQVFTYKOSTTTHOK 50  
DB 9 GSGVWVDPDEAMLDGEVVEANGOEI--KVNCCQKTVSPSPKORDVVLTK----VWAK 62  
QY 51 VTAMHPNEE-----GVDDMASLTBLHGGSIMVNLFPQRYKNOQWTYIGSLASVNYQPI 106  
DB 63 VNVVHPDPEFPPELGVDMDTKLAVLHEPGVLLNKARVNAEITYTGNTLLVNVNPFKRI 122  
QY 107 AGLYEPATMEQYSRRHLGELPHIIFAIANECYRCRLMKRHDNOCILKGSAGAKTSTKL 166  
DB 123 PHLYGNEIMQYKGTDBGELSPHPFAVADSAYRKMINEGVSQAILVSGSAGAKTSTTKM 182  
QY 167 ILKFLSVISQOSELSEKTSCEVERAILESSPIMEAFGNAKTVYNNSSRFQKPVQANI 226  
DB 183 LMQYLAVMGKA-----ESEGRSVGQOVLESNPVLEAFGNAKTVYNNSSRFQKPVYIOP 237  
QY 227 CQKGNIOGRIIVDCILSSQNRVVRQNGERNYHIFYALLAGLEHREBEFYLTSPENHY 286  
DB 238 NHMGRISGAARITYLL-ERSRVCQSDPERNYHCFYMLCAAP-QETERYQKGFSTFHY 295  
QY 287 LNSGCEVEDTISDOESFEVITAMDVQPSKEEVRSRLAGILHGNIEFI---TAG 343  
DB 296 LNSGNCALAIIDSKEYLATRKAMDVGISPEQDAIFRVVAAILHGNIEFAKSESD 355  
QY 344 GAQV---SFTYALGRSHELLGLDPTQDALTQSWFLRGEELTPLVNOQAVDSRSLA 400  
DB 356 GAEPKDKSRPHLKVAKLFMCDEKALENSLCNRVMTGDESITKPLDPSALSRLDAL 415  
QY 401 MALYACCFEVIKINSRIGNEDFK-SIGILDPGENFENVHFGQFNINANEKLOEY 459  
DB 416 KIYYSKLFDMLVTKINNSIQDSSSKYIIGVDIYGESFETKNSFROFCINTNEKLOOH 475  
QY 460 FNKHFELQLEYSREGLVWEDIDWIDNGECLDLEKUL-GLLALINBSHPQATDSTL 518  
DB 476 FNGHVPRMEQBEYTKEDIDSYIEFDINQVLDLIEKKBGIIIALDDEACMFPRSTHDL 535  
QY 519 LEKLHSGHANNHFYVKRVAVNNFGVKGAVGEVQYDVGILKKNRDTFRDDLNLRESR 578  
DB 536 AEKLYQTFGSHKFTPKLARTDFTICHYAGDYQTELFDKKNDYVGEHQSLSMNSD 595  
QY 579 FDIYDLFEHVSSRNQDTLKCGSKHRFPVSSQFKVDSLHSLMATLSSNPFVRCIKP 638  
DB 596 CSFVSSLF---PKSREBSK---SSKFSISIGQFK-QOQSLLETLNTEPHYIRCVKP 647  
QY 639 NMQKMPDQDAVVLNQLRYSGLMETVRIKAGAVARPFQDFYKRYKVLNRNALPE-- 696  
DB 648 NNVLKPEIFENNVVNLHQLRCGVMEAIRISCAGYPTKRPNEFLTRFIL-----APEAT 702  
QY 697 ----DVGRKCTSLQLYDASNSEWOLGKTVFLRESLEOKLEKRREBEVSHAAMVIRAHV 752  
DB 703 ERSFDEVDACKLLARVDLKG--FOIGKTKVFLRAGOMELDAHRAVLSHSARIIORKV 760  
QY 753 LGFLARKY-----RKY 764  
DB 761 ITYLSRRKYLLLOSASTETQAFCRGHIAVQFATRREASVRIQOARTYICQTAFAKL 820  
QY 765 LYCVIILQKNYRAF-----LLRRF 784  
DB 821 CASAIISQSLRAMARVEFOYRTKKAATIIQASLKPHIDDKLSFQSIRKLCRRKY 880  
QY 785 LHLKKAIVFOQLRGQIARVYROLAERK---QEEKKQEEB-----EKKK 830  
DB 881 LRTTKAAITTCQGMRYVAHRELRKLMMAKETGALQDAKTKLEKEVEBELTSCLEKQM 940  
QY 831 REEERERERERE-----AELRAQEEETRKQOELALOKSOKE-AELTREL 877

DB 941 RMELDQVKTQEVEDLRSALNDMKLOLGETQVTKSEIILKLO--SALQDMOLFEBELAKEL 998  
QY 878 EKQK-----ENKQVEEIL-RLKEKIEDL-----ORKK-----E 904  
DB 999 EMTNLLAENEDLXOLVSLQKIDESDKYBETSKLSBERVKQEVPIIDGVIIKLEAE 1058  
QY 905 QOELSLTEASLOK-----LOERRDQELRLBEERACAAQAFLESINPDE-- 948  
DB 1059 NQKLALVSTLEKKIDSLDRKHDDLVDLLERKIDETEKKYEBASKLCEERLKQVVDTEKX 1118  
QY 949 IDECVNTERLSGGSFESSLEASACEKPRFNSQPYPEEYVDEGFEADDDAF----- 1003  
DB 1119 YEEASHLCEERLKQVVDTEKLE-----LKTSMQRLBEKVSD-MEAEKIIIRQOAL 1169  
QY 1004 -----KQSP-----NPEHQHSD 1016  
DB 1170 RNSASRKMSPOKSLDLFVPMYLFQGVENGHHE 1201

Search completed: July 14, 2003, 18:21:37  
Job time : 122.43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:02:16 ; Search time 73.051 Seconds  
(without alignments)  
3843.329 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936

Sequence: 1 MSQPLLPASAEKTRKTRAL.....ISGTPSTWVHGHTSSSSVV 2107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A.GeneSeq\_101002.\*

1:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23:	/SID2/gcgcdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10936	100.0	2107	AAE11889	Angiogenesis assoc
2	9926.5	90.8	1980	ABG61706	Human cadherin-11k
3	9918.5	90.7	1980	ABG61702	Cadherin-11k asym
4	8173	74.7	1572	AAAB40315	Human ORFX ORF79 p
5	7976	72.9	1534	ABG61676	Cadherin-11k asym
6	7976	72.9	1534	ABG61676	Cadherin-11k asym
7	7976	72.9	1534	ABG61680	Cadherin-11k asym
8	6784.5	62.0	1352	AAAB36527	Human CLASP-2 prot
9	6784.5	62.0	1352	AAAB36528	Human CLASP-2A pro
10	6784.5	62.0	1353	ABG61670	Cadherin-11k asym

11	6784.5	62.0	1353	23	ABG61671
12	6784.5	62.0	1353	23	ABG61672
13	6784.5	62.0	1353	23	ABG61686
14	6782.5	62.0	1353	23	ABG61679
15	6782.5	62.0	1353	23	ABG61687
16	6149	56.2	1235	23	ABG61674
17	6105.5	55.8	2008	22	AAU04090
18	6105.5	55.8	2008	23	ABG61703
19	6022.5	55.1	1214	21	ABG61677
20	5930.5	54.2	1194	23	AAAB36530
21	5930.5	54.2	1194	21	AAAB36532
22	5930.5	54.2	1194	23	ABG61694
23	5930.5	54.2	1194	23	ABG61695
24	5930.5	54.2	1194	23	ABG61696
25	5930.5	54.2	1194	23	ABG61697
26	5930.5	54.2	1194	23	ABG61698
27	5930.5	54.2	1194	23	ABG61699
28	5930.5	54.2	1194	23	ABG61700
29	5930.5	54.2	1194	23	ABG61701
30	5928.5	54.2	1194	21	AAAB36536
31	5924.5	54.2	1194	21	AAAB36529
32	5924.5	54.2	1194	21	AAAB36535
33	5923.5	54.2	1194	21	AAAB36531
34	5922	54.2	1157	22	AAAB9502
35	5916.5	54.1	1194	22	AAAB9503
36	5753.5	52.6	1194	21	AAAB36533
37	5753.5	52.6	1194	21	AAAB36534
38	5560	50.8	2180	23	ABG61708
39	5560	50.8	2180	23	AAU1752
40	5489	50.2	1177	22	AAAB6406
41	4682.5	42.8	955	23	ABG61673
42	3689.5	33.7	1215	23	ABG61696
43	3608.5	33.0	1215	23	ABG61682
44	3595.5	32.9	1214	22	AAU04091
45	3442.5	31.5	1289	21	AAAB36521

## ALIGNMENTS

RESULT 1  
AAE11889  
ID AAE11889 standard; Protein; 2107 AA.

AC AAE11889;  
DT 18-DEC-2001 (first entry)  
DE Angiogenesis associated human thyroid regulated transcript (HTRG).

KW Angiogenesis associated protein; AAP; cytosolic; cardiac; gene therapy;  
KW ophthalmological; vulnary; myocardial infarction; macular degeneration;  
KW diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;  
KW rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;  
KW cancer; therapeutic; diagnostic; human; thyroid regulated transcript;  
KW TRG.

OS Homo sapiens.  
XX  
XX WO200170808-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 22-MAR-2001; 2001WO-US09609.  
XX  
XX PR 22-MAR-2000; 2000US-191134P.  
XX  
XX PA (CURA-) CURAGEN CORP.  
XX (GENTH) GENTECH INC.  
XX Rastelli LK, Gerritsen M;  
XX WPI; 2001-602775/68.  
XX

Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Human cadherin-11k  
Human cadherin-11k  
Preliminary CLASP-  
Preliminary CLASP-  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Preliminary CLASP-  
Preliminary CLASP-  
Preliminary CLASP-  
CLASP-3 protein fr  
CLASP-3 protein fr  
Preliminary CLASP-  
Preliminary CLASP-  
Human cadherin-11k  
Human cadherin-11k  
Amino acid sequenc  
Cadherin-11k asym  
Cadherin-11k asym  
Cadherin-11k asym  
Human CLASP-4 rela  
Mouse CLASP-1 prot

DR N-PSDB; AAD19118.  
XX Novel angiogenesis associated polypeptides and polynucleotides encoding  
PT the polypeptides, useful for modulating angiogenesis and for treating  
PT tumors and cancers -  
XX  
XX claim 1; Page 21-28; 159pp; English.  
XX  
XX The invention relates to angiogenesis associated proteins (AAP) and their  
CC corresponding cDNA molecules, which are useful for modulating  
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting  
CC wound healing, for example after organ transplantation, and in the  
CC treatment of tumours, myocardial infarction, cancers, diabetic  
CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
CC AAP proteins and DNA's are useful in potential prophylactic and  
CC therapeutic applications implicated in a variety of disorders including  
CC those related to angiogenesis, and also in diagnostic applications.  
CC AAP cDNA is also useful in gene therapy. The invention also relates to  
CC a method for screening a tissue sample for tumourigenic potential. AAP  
CC proteins are used to screen drugs or compounds that modulate AAP activity  
CC or expression as well as treating disorders characterised by insufficient  
CC or excessive production of AAP or production of AAP forms that have  
CC decreased or aberrant activity compared to the wild type protein, or  
CC modulate biological function that involve AAP. The present sequence  
CC is human thyroid regulated transcript (hTRG) protein which is an  
CC angiogenesis associated protein (AAP) of the invention. Human TRG is  
CC upregulated in the in vitro model of angiogenesis and is likely to be  
CC involved in signal transduction between receptors and kinases. Modulation  
CC of hTRG is useful to treat diseases related to thyroid stimulating  
CC hormone (TSH) imbalance.  
XX  
XX Sequence 2107 AA,  
Query Match 100.0%; Score 10936; DB 22; Length 2107;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 DNAAFSATIRQDSNKLSDNDMLKLADFRKPEKAKLPIVILGNDITIDNVSDPENYV 600  
DB 541 DNAAFSATIRQDSNKLSDNDMLKLADFRKPEKAKLPIVILGNDITIDNVSDPENYV 600  
QY 601 NSSYIPTOFENCSTPIITFEVEEFVPCIPHTOPTITITNHLVYPRKLYVDSQSKPAK 660  
DB 601 NSSYIPTOFENCSTPIITFEVEEFVPCIPHTOPTITITNHLVYPRKLYVDSQSKPAK 660  
QY 661 ARNIAICIEFKDDEDEQPLKCIYRPGGVPFTSAAVALHHNQNEPFYDEIKIEIPT 720  
DB 661 ARNIAICIEFKDDEDEQPLKCIYRPGGVPFTSAAVALHHNQNEPFYDEIKIEIPT 720  
QY 721 QHEKHHLILTFEHSVCSNNSKSGTCKRDVVEVTOGVSWPLKDGRTVTSQIIPVAN 780  
DB 721 QHEKHHLILTFEHSVCSNNSKSGTCKRDVVEVTOGVSWPLKDGRTVTSQIIPVAN 780  
QY 781 LPSGYLGOELGMDGHRYPETIKVNDGKPLKISHLVSTVYTDQHLNPFQYCKTES 840  
DB 781 LPSGYLGOELGMDGHRYPETIKVNDGKPLKISHLVSTVYTDQHLNPFQYCKTES 840  
QY 841 GAQALGNELVYKLSLHAMEGHVMIAPPTILNQLFRVLTRATQBEVAVNVTYVLIHVA 900  
DB 841 GAQALGNELVYKLSLHAMEGHVMIAPPTILNQLFRVLTRATQBEVAVNVTYVLIHVA 900  
QY 901 QCHEEGLSHLSRYKVAIYKAPVYASGYTVHEBLTYSMTTILKPSADFLTSNKLKYS 960  
DB 901 QCHEEGLSHLSRYKVAIYKAPVYASGYTVHEBLTYSMTTILKPSADFLTSNKLKYS 960  
QY 961 WFFPVLTKSMQHLIENSKVYLLNORFPASVYHVAETVNMMLPHITOKFRDPEASK 1020  
DB 961 WFFPVLTKSMQHLIENSKVYLLNORFPASVYHVAETVNMMLPHITOKFRDPEASK 1020  
QY 1021 NANSLAVFIKRCFTFMDRGFVFKQINNYISCFAPADPKTLPEYKFEELRVVCNHEHYIP 1080  
DB 1021 NANSLAVFIKRCFTFMDRGFVFKQINNYISCFAPADPKTLPEYKFEELRVVCNHEHYIP 1080  
QY 1081 LNLPMFPGKRIQRYQDILQDLSLTDFCRNHFVGLLREVTALQDFREYRLAIVYL 1140  
DB 1081 LNLPMFPGKRIQRYQDILQDLSLTDFCRNHFVGLLREVTALQDFREYRLAIVYL 1140  
QY 1141 KNLILKHSFDDRYARSHOARATLYLPLFGLLINVORINRVDSPPPVNAGMTVQDES 1200  
DB 1141 KNLILKHSFDDRYARSHOARATLYLPLFGLLINVORINRVDSPPPVNAGMTVQDES 1200  
QY 1201 LALPAVNPVLTVPQKSTLDNSLHKDLGASIGIASPYTSTSPINNSVNNADSGSLISTD 1260  
DB 1201 LALPAVNPVLTVPQKSTLDNSLHKDLGASIGIASPYTSTSPINNSVNNADSGSLISTD 1260  
QY 1261 SGNSLPERNSEKSNLDRHQOSTLGNVVRCDKLDQSEIKSLMCFYILKSMDDDLAF 1320  
DB 1261 SGNSLPERNSEKSNLDRHQOSTLGNVVRCDKLDQSEIKSLMCFYILKSMDDDLAF 1320  
QY 1321 TYNNKASTSELMDFPTISEVCLHOFQYMGKRIARTGMMHARLQOLGSLDNLSTFNHSYG 1380  
DB 1321 TYNNKASTSELMDFPTISEVCLHOFQYMGKRIARTGMMHARLQOLGSLDNLSTFNHSYG 1380  
QY 1381 HSDADVLQSLLEANIATEVCLTALDLSLFTLAFKQNLADHGNPIMKVFVYLCEL 1440  
DB 1381 HSDADVLQSLLEANIATEVCLTALDLSLFTLAFKQNLADHGNPIMKVFVYLCEL 1440  
QY 1441 QHGOSETALKVFTLRSILYFPESTFYEGRADMCALCYELKCNLSIRTEASQL 1500  
DB 1441 QHGOSETALKVFTLRSILYFPESTFYEGRADMCALCYELKCNLSIRTEASQL 1500  
QY 1501 LYFLMRNPFYDGKSPFRTHQVLIISVQLADVAGIGTRFOOSLSTINNCAUSDRLI 1560  
DB 1501 LYFLMRNPFYDGKSPFRTHQVLIISVQLADVAGIGTRFOOSLSTINNCAUSDRLI 1560  
QY 1561 KHTSFSSDYKDLTKIRITVLAATAQMKHENDPEMLVLDQSLAKSYASTPELRKTWLD 1620  
DB 1561 KHTSFSSDYKDLTKIRITVLAATAQMKHENDPEMLVLDQSLAKSYASTPELRKTWLD 1620  
QY 1621 MARIHVKGDLSEAAACVHTALVAEYLTRKAVQMBEPLLPHSHSACLRBSRGVFRQ 1680

```

Db      1621 MARHVKGDLSSEAMCYHVALVALEYLTRKEAVQWEPPLPHSHSACLRSRGVFQ 1680
Qy      1681 GCTAFVITPNIDEASNMEDVGMODVFNEDVLMELBCADGLKAKAREYELIADYKL 1740
Db      1681 GCTAFVITPNIDEASNMEDVGMODVFNEDVLMELBCADGLKAKAREYELIADYKL 1740
Qy      1681 GCTAFVITPNIDEASNMEDVGMODVFNEDVLMELBCADGLKAKAREYELIADYKL 1740
Db      1681 GCTAFVITPNIDEASNMEDVGMODVFNEDVLMELBCADGLKAKAREYELIADYKL 1740
Qy      1741 IIPYHKRRDPERLAHLVDTLHRAVSKVTEVMHSGRLLGTYFRVAFGQAAYOFTDSE 1800
Db      1741 IIPYHKRRDPERLAHLVDTLHRAVSKVTEVMHSGRLLGTYFRVAFGQAAYOFTDSE 1800
Qy      1801 TVEGHPFEDDEKEYIYKPKLTPLEISIQRLKLYSDKFGSENMVIMIDSGKNPKDLD 1860
Db      1801 TVEGHPFEDDEKEYIYKPKLTPLEISIQRLKLYSDKFGSENMVIMIDSGKNPKDLD 1860
Qy      1801 TVEGHPFEDDEKEYIYKPKLTPLEISIQRLKLYSDKFGSENMVIMIDSGKNPKDLD 1860
Db      1861 SKYAYIQVTHVLPFPEKELOERKTEPERSHNIRRMPEKPFQTQKRGVBEQCKRRT 1920
Qy      1861 SKYAYIQVTHVLPFPEKELOERKTEPERSHNIRRMPEKPFQTQKRGVBEQCKRRT 1920
Db      1921 ILTALHCPYVKRRIIPVYQHTDLNPIEVAIDENSKYAEELRQLCSAEVDMIKQLKLT 1980
Qy      1921 ILTALHCPYVKRRIIPVYQHTDLNPIEVAIDENSKYAEELRQLCSAEVDMIKQLKLT 1980
Db      1921 ILTALHCPYVKRRIIPVYQHTDLNPIEVAIDENSKYAEELRQLCSAEVDMIKQLKLT 1980
Qy      1981 QGSVSVQVAGPLAYARAFLDNTNTRYDGNKYLKKEYFRQVEACGALAVNERLKE 2040
Db      1981 QGSVSVQVAGPLAYARAFLDNTNTRYDGNKYLKKEYFRQVEACGALAVNERLKE 2040
Qy      2041 DPLEYOEKANYREMAKELSEIMHEQICPLEEKTSLVNSLIHFAISGTPSTVWHGM 2100
Db      2041 DPLEYOEKANYREMAKELSEIMHEQICPLEEKTSLVNSLIHFAISGTPSTVWHGM 2100
Qy      2101 TSSSSSV 2107
Db      2101 TSSSSSV 2107

```

```

PT      New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT      an immune response, and for treating multiple sclerosis, rheumatoid
PT      arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT      and sepsis
PS      Disclosure, Figure 13; 245pp; English.
XX
CC      The invention relates to an isolated polypeptide (I) comprising an amino
CC      acid sequence that has 90 % sequence identity to one of the human
CC      cadherin-like asymmetry protein(s) (CLASP)-2 (CLASP-2A, 2B, 2C, 2E)
CC      sequences (PS). (I) is useful for identifying a compound or agent that
CC      binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
CC      detecting a CLASP-2 polypeptide in a sample. (II) is useful for
CC      inhibiting an immune response in a subject. A pharmaceutical composition
CC      comprising a nucleic acid encoding (i), or (ii) is useful for preventing
CC      or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
CC      the autoimmune disease is caused or exacerbated by increased activity
CC      of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
CC      detection or inhibition of CLASP-2 expression (e.g., antisense or
CC      ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
CC      polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
CC      antibodies or are used as therapeutic polypeptides. The CLASP-2
CC      polynucleotide or fragments can be used in diagnostics (e.g., as probes
CC      for CLASP-2 expression), as a lymphocyte marker and for therapeutic
CC      purposes. CLASP-2 polynucleotides can construct transgenic and knockout
CC      animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
CC      polynucleotides can screen for CLASP-2 agonists and antagonists.
CC      CLASP-2 polypeptides or polynucleotides can treat deficiencies or
CC      disorders of the immune system, by activating or inhibiting the
CC      activation, differentiation of immune cells and can treat or detect
CC      deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
CC      or polynucleotides can increase differentiation and proliferation of
CC      haematopoietic cells, including the pluripotent stem cells to treat those
CC      disorders associated with a decrease in certain (or many) types of
CC      haematopoietic cells e.g., immunologic deficiency syndromes including
CC      blood protein disorders (e.g., agammaglobulinemia),
CC      dysgammaglobulinemia, ataxia telangiectasia, common variable
CC      immunodeficiency, DiGeorge syndrome, lymphopenia, thrombocytopenia, or
CC      haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
CC      detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
CC      Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
CC      endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
CC      inflammation. CLASP-2 can be used to treat anaphylaxis or
CC      hypersensitivity to an antigenic molecules, organ rejection or graft-
CC      versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
CC      represent cadherin-like asymmetry protein (CLASP) sequences of the
CC      invention.
XX
SQ      Sequence 1980 AA;
Qy      Query Match 90.8%; Score 9926.5; DB 23; Length 1980;
Db      Best Local Similarity 93.8%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;

```

QY	371	CNDLS.FNLOCCAENEBCGTTNVEPFTYLSLFDIKNRKISADPHDNLNHSVCOMAT	430
Db	300	CNDLS.FNLOCCAENEBCGTTNVEPFTYLSLFDIKNRKISADPHDNLNHSVCOMAT	359
QY	431	TSPALMNGSGPPTOSALRGILHEAAMQVPEKQIFSVTCPHDIFLVARIEKYLOGSIYHC	490
Db	360	TSPALMNGSG-QSPVLKGLIHEAAMQVPEKQIFSVTCPHDIFLVARIEKYLOGSIYHC	418
QY	491	AEPYKSSSPSSKYAQKVLKNAQACORIGQYMPRMAARTL.FKQASGLDKNAFSAIY	550
Db	419	AEPYKSSSPSSKYAQKVLKNAQACORIGQYMPRMAARTL.FKQASGLDKNAFSAIY	478
QY	551	RODSNKL.SNDDMLKLLADPRKEPKMAKLPVILGNDITLDNVSSDPFVNVNVSSTY.PTQOF	610
Db	479	RODSNKL.SNDDMLKLLADPRKEPKMAKLPVILGNDITLDNVSSDPFVNVNVSSTY.PTQOF	538
QY	611	ETCSKTPITFEVEBEFVPCIPKHTOPYTYTNMLVYVYKLYKSQKSPAKANINACIEF	670
Db	539	ETCSKTPITFEVEBEFVPCIPKHTOPYTYTNMLVYVYKLYKSQKSPAKANINACIEF	598
QY	671	KDSDEDSQPLKCIYRPGCGPVFTSAPAAVLIHHNBNPFYDEIKTELPTOLHEGHILL	730
Db	599	KDSDEDSQPLKCIYRPGCGPVFTSAPAAVLIHHNBNPFYDEIKTELPTOLHEGHILL	658
QY	731	TFPHVSCDNSSKGSYTKKQDVVEYQVGSYMLPLLKQGVVTSBQHI.PVSAIPLPSGLGYOE	790
Db	659	TFPHVSCDNSSKGSYTKKQDVVEYQVGSYMLPLLKQGVVTSBQHI.PVSAIPLPSGLGYOE	718
QY	791	LGMGHHYGEIEMVNDGGRPLKLSIHLVSTVYTOOHLNPFQYQCKTESGQALGMBLY	850
Db	719	LGMGHHYGEIEMVNDGGRPLKLSIHLVSTVYTOOHLNPFQYQCKTESGQALGMBLY	778
QY	851	KYLKSLIHMHEGVMIAFLPTLIHQ.FRYVLTRATQEEVAVNVTWVLIHHVACQHEGLES	910
Db	779	KYLKSLIHMHEGVMIAFLPTLIHQ.FRYVLTRATQEEVAVNVTWVLIHHVACQHEGLES	838
QY	911	LRSYKVAIKAEPPYVASEKYTVHEELIYKSMITILKPSADFLTSLNKLKYSWFFDVLLKS	970
Db	839	LRSYKVAIKAEPPYVASEKYTVHEELIYKSMITILKPSADFLTSLNKLKYSWFFDVLLKS	898
QY	971	MAQHLIENSKVYKLLNORPASVYHNAVEVNVNMLMPHITOKRDNDEASKANHSIAFI	1030
Db	899	MAQHLIENSKVYKLLNORPASVYHNAVEVNVNMLMPHITOKRDNDEASKANHSIAFI	958
QY	1031	KKCFETPMDRGFVFKQINNYSICFPAEGDPKTLIEYKFEELRVVCNHEHYIPLNLPMPFGKG	1090
Db	959	KKCFETPMDRGFVFKQINNYSICFPAEGDPKTLIEYKFEELRVVCNHEHYIPLNLPMPFGKG	1018
QY	1091	RIQORVODLODVSILNDEPCRNHFLVGLLREVEGTLQOERREVRLLAISVLYKMLIKHSFD	1150
Db	1019	RIQORVODLODVSILNDEPCRNHFLVGLLREVEGTLQOERREVRLLAISVLYKMLIKHSFD	1078
QY	1151	DRYASRSHQARIATATYLP.FGLLIENVORINRDVSPPVNAGMTVKQESLALPAPNVLY	1210
Db	1079	DRYASRSHQARIATATYLP.FGLLIENVORINRDVSPPVNAGMTVKQESLALPAPNVLY	1138
QY	1211	TPQKSTLDNLSLKHOLLGALISGIASPYTTSTYNINSVRNADSRGSLISTDSGNSLPERNS	1270
Db	1139	TPQKSTLDNLSLKHOLLGALISGIASPYTTSTYNINSVRNADSRGSLISTDSGNSLPERNS	1198
QY	1271	EKSNLSLDKHOQSSSTLGNSVVRCDKLDOSEIKSLMCPYLYILKSMDDDLFTYMNKASRSE	1330
Db	1199	EKSNLSLDKHOQSSSTLGNSVVRCDKLDOSEIKSLMCPYLYILKSMDDDLFTYMNKASRSE	1258
QY	1331	IMDPFTTISEVCLHQFOYMGKRYIA-----RTGMHARLQOLG	1367
Db	1259	IMDPFTTISEVCLHQFOYMGKRYIA-----RTGMHARLQOLG	1318
QY	1368	SLDNLSTLPHHSGHSDADVLHQSLLEANINATEVCITALDITSLFTLARKNQLADHGNNP	1422
Db	1319	SLDNLSTLPHHSGHSDADVLHQSLLEANINATEVCITALDITSLFTLARKNQLADHGNNP	1378

QY	1428	LMKKVFDVYLGFLOKHQSETLKNVPALNSLIYKPPSTPEEGADNCAALCYEILKCN	1467
Db	1379	LMKKVFDVYLGFLOKHQSETLKNVPALNSLIYKPPSTPEEGADNCAALCYEILKCN	1438
QY	1488	SKLSIRFEASQQLLYFLMRNNFDYTGKSKFVRTHAOYIISVSQLIADVVGIGTRFOOSL	1547
Db	1439	SKLSIRFEASQQLLYFLMRNNFDYTGKSKFVRTHAOYIISVSQLIADVVGIGTRFOOSL	1498
QY	1548	SIINNCANSRLIHTSFSSDVKOLTKRIRFTVLMATQMKHEHNDPEMLVDLOYSLAKSY	1607
Db	1499	SIINNCANSRLIHTSFSSDVKOLTKRIRFTVLMATQMKHEHNDPEMLVDLOYSLAKSY	1558
QY	1608	ASTBELRKTWLDSMARLIHVKNKGDISEAMCVHTALVAEYLTKENAVQNEPRLPHSHS	1667
Db	1559	ASTBELRKTWLDSMARLIHVKNKGDISEAMCVHTALVAEYLTK-----	1603
QY	1668	ACLRRSRGVFRQCGTAFRVITPNIIDEASMMEDVGMODVFNEDVIMELLEQCADGLMK	1727
Db	1604	-----GVRRQCGTAFRVITPNIIDEASMMEDVGMODVFNEDVIMELLEQCADGLMK	1655
QY	1728	AERYELIADIYKLIIPYEKRRDPERLAHLYDTLHRAYSKYTEVMHSGRLLGTYPVAF	1787
Db	1656	AERYELIADIYKLIIPYEKRRD-----	1678
QY	1788	FGQAQVQFTDSEIDVGEFPEDEGCKEYIYKBPCLTPLSEISORLLKLYSDPKSSENVKM	1847
Db	1679	-----FPEDEGCKEYIYKBPCLTPLSEISORLLKLYSDPKSSENVKM	1720
QY	1848	IODSGKNPKDLDSKVAIYQVTHVIPFDEKELOERKTEPFERSNNIRFMPEMPFTQTGK	1907
Db	1721	IODSGKNPKDLDSKVAIYQVTHVIPFDEKELOERKTEPFERSNNIRFMPEMPFTQTGK	1780
QY	1908	RQGVBEQCKRRITLILAIHCFPPYKGIIPWYQHTDNLNPIEVALIDEMSKKVAELROLCS	1967
Db	1781	RQGVBEQCKRRITLILAIHCFPPYKGIIPWYQHTDNLNPIEVALIDEMSKKVAELROLCS	1840
QY	1968	SAEVDMLTQLKLOGSSVSQVNAAGPLAYARAFLDITTKRYPDNKKVYLKEVFQPEAC	2027
Db	1841	SAEVDMLTQLKLOGSSVSQVNAAGPLAYARAFLDITTKRYPDNKKVYLKEVFQPEAC	1900
QY	2028	GOALAVNERLIKEDOLEYOEEMKANYREMAKELSEIMHEQICPLEEKTSVLPNSLHIFNA	2087
Db	1901	GOALAVNERLIKEDOLEYOEEMKANYREMAKELSEIMHEQICPLEEKTSVLPNSLHIFNA	1960
QY	2088	ISGTPTSTMVHGMTSSSSSV 2107	
Db	1961	ISGTPTSTMVHGMTSSSSSV 1980	
RESULT 3			
ABG61702			
ID	ABG61702	standard; Protein; 1980 AA.	
XX	XX		
AC	ABG61702;		
DT	13-AUG-2002	(first entry)	
XX	XX		
DE	Cadherin-like asymmetry protein (CLASP) encoded by exon 2.		
XX	XX		
KM	Human; autoimmune disease; haematopoietic disorder; Digesorge syndrome;		
KM	blood protein disorder; agammaglobulinaemia; dysgammaglobulinaemia;		
KM	ataxia telangiectasia; common variable immunodeficiency; lymphopenia;		
KM	chromocytopenia; haemoglobinuria; Addison's disease; Grave's disease;		
KM	haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;		
KM	endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;		
KM	autoimmune pulmonary inflammation; organ rejection; inflammation;		
XX	CLASP.		
OS	Homo sapiens.		
FN	WO200231117-A2.		
XX	18-APR-2002.		
PD			



XX 15-OCT-2001; 2001WO-US32202.  
 PF 13-OCT-2000; 2000US-0687837.  
 XX (ARBO-) ARBOR VITAE CORP.  
 XX (GARW/) GARMAN J D.  
 PA (CAND/) CANDIA A F.  
 XX Lu PS;  
 PI WPI, 2002-416861/44.  
 DR N-PSDB; ABR85003.  
 XX New human cadherin-like asymmetry protein(s) (CLASP)-2 for mediating  
 PT an immune response, and for treating multiple sclerosis, rheumatoid  
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,  
 PT and sepsis -  
 XX Example 1; Figure 11, 245pp; English.  
 PS The invention relates to an isolated polypeptide (I) comprising an amino  
 CC acid sequence that has 90 % sequence identity to one of the human  
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2c, 2e)  
 CC sequences (PS). (I) is useful for identifying a compound or agent that  
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 CC inhibiting a immune response in a subject. A pharmaceutical composition  
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing  
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 CC the autoimmune disease is caused or exacerbated by increased activity  
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or  
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2  
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2  
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or  
 CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, dysgerge syndrome, lymphopenia, thrombocytopenia, or  
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or  
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,  
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,  
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary  
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or  
 CC hypersensitivity to an antigenic molecule, organ rejection or graft-  
 CC versus-host disease (GVHD) and inflammation. ABB61670-ABB61708  
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the  
 CC invention.  
 CC Sequence 1980 AA;  
 SQ

Query Match 90.7%; Score 9918.5; DB 23; Length 1980;  
 Beel Local Similarity 93.7%; Pred. No. 0;  
 Matches 1931; Conservative 8; Mismatches 16; Indels 105; Gaps 6;

Qy 73 MLLPFDQATIALRQGRYICSTVPAKAEENQSLPTECIKTYSWHLVNYKEDS 132  
 Db 1 MLLPFDQATIALRQGRYICSTVPAKAEENQSLPTECIKTYSWHLVNYKEDS 60

Qy 133 GEFROLFNKRVKLDKLPVHYVEVDEEVDKDDAASLGQKGIYTKHGLYKGNMNSAISV 192

Db 61 GEFROLFNKRVKLDKLPVHYVEVDEEVDKDDAASLGQKGIYTKHGLYKGNMNSAISV 120  
 Qy 193 TMRSPFRFPFHLIOLGDSGYKPEFLKD--LQKEPGSI.FLPELVGVSPNNKVRPAPFL 250  
 Db 121 TMRSPFRFPFHLIOLGDSGYKPEFLKD--LQKEPGSI.FLPELVGVSPNNKVRPAPFL 179  
 Qy 251 KMDDKSSYLLAADSEVEEMWITTLNKLQLNFEAAMQEKNGDSHEDEOSKLGSGSG 310  
 Db 180 KMDDKSSYLLAADSEVEEMWITTLNKLQLNFEAAMQEKNGDSHEDEOSKLGSGSG 239  
 Qy 311 LDSTYLPBLAKAAREAEITLKBSRYKFLYLDPPAQKLDPSAEPVKSPEEKFGRIIVK 370  
 Db 240 LDSTYLPBLAKAAREAEITLKBSRYKFLYLDPPAQKLDPSAEPVKSPEEKFGRIIVK 299  
 Qy 371 CNDLSFNLQCCVAENEBEPTTNVPEFVTLSPDIKYNRKISADPHVNLNPSVQOMAT 430  
 Db 300 CNDLSFNLQCCVAENEBEPTTNVPEFVTLSPDIKYNRKISADPHVNLNPSVQOMAT 359  
 Qy 431 TSPALMNGSGPETOALRGILHEAAMQYKXGIFSVTCPHPDIFLVANIBKVLQGSITHC 490  
 Db 360 TSPALMNGSG-QSPEVLKGLIHEAAMQYKXGIFSVTCPHPDIFLVANIBKVLQGSITHC 418  
 Qy 491 AEPYKSSDSSKVAQKVLAKAKQACQRLGQRMPPAMAARTLFDKASGNDLKNARFSAIY 550  
 Db 419 AEPYKSSDSSKVAQKVLAKAKQACQRLGQRMPPAMAARTLFDKASGNDLKNARFSAIY 478  
 Qy 551 RODSNKLSNDMLKLLAPFRPEKAKLPVLGNLNDITDNDVSSPFPYVNSSYPTQOF 610  
 Db 479 RODSNKLSNDMLKLLAPFRPEKAKLPVLGNLNDITDNDVSSPFPYVNSSYPTQOF 538  
 Qy 611 ETCSKTPTTFEVEBEFPCIPKHTQPYTITNHLVYPPYKLYDSQSKPAKARNIACIEF 670  
 Db 539 ETCSKTPTTFEVEBEFPCIPKHTQPYTITNHLVYPPYKLYDSQSKPAKARNIACIEF 598  
 Qy 671 KDSDEDSQPLKCTYGRGCGVFTRSAVAALVHHQNEPDEIKIELPTQLHEKHILL 730  
 Db 599 KDSDEDSQPLKCTYGRGCGVFTRSAVAALVHHQNEPDEIKIELPTQLHEKHILL 658  
 Qy 731 TPFHNSCNSSKSGSKRDKDVETQVGSWMLPLDKGRVVTSEOHIPVABNIPSGLVQOE 790  
 Db 659 TPFHNSCNSSKSGSKRDKDVETQVGSWMLPLDKGRVVTSEOHIPVABNIPSGLVQOE 718  
 Qy 791 LGMGRHYPELKWMDGGRPLKISTHLVSTVYTDQHLNFFQYQCKTESGAQALGNELV 850  
 Db 719 LGMGRHYPELKWMDGGRPLKISTHLVSTVYTDQHLNFFQYQCKTESGAQALGNELV 778  
 Qy 851 KTLKSLHAMEGHVMIAPLPTILNQLFRVLTATQDEAVANVTYVLIHVVAQCHEGLES 910  
 Db 779 KTLKSLHAMEGHVMIAPLPTILNQLFRVLTATQDEAVANVTYVLIHVVAQCHEGLES 838  
 Qy 911 LRSYKVAKYKPEPYVASYKYVHELTSMITLKPSADPLTSNKLKYSWFFPVILKS 970  
 Db 839 LRSYKVAKYKPEPYVASYKYVHELTSMITLKPSADPLTSNKLKYSWFFPVILKS 898  
 Qy 971 MAOHLENSKYVLENGRFPASYNHAAVETVNMMLPHITQKRDNPBASKUANHSLAVFI 1030  
 Db 899 MAOHLENSKYVLENGRFPASYNHAAVETVNMMLPHITQKRDNPBASKUANHSLAVFI 958  
 Qy 1031 KRCEFTMDRGVFNQINNYISCPAFGDPKTLPEYKFEFLAVVCNHEHYIPLNLPMPFGK 1090  
 Db 959 KRCEFTMDRGVFNQINNYISCPAFGDPKTLPEYKFEFLAVVCNHEHYIPLNLPMPFGK 1018  
 Qy 1091 RIORQDOLDYSLDFECRNHPLVGLLRVGTALQEFREVRLLAISVKNLLIKHSFD 1150  
 Db 1019 RIORQDOLDYSLDFECRNHPLVGLLRVGTALQEFREVRLLAISVKNLLIKHSFD 1078  
 Qy 1151 DRYASRSHQARIATLPLFGILLIENVQRIINVRDVSPPVNAQMTVKDESLALPAVNPVY 1210  
 Db 1079 DRYASRSHQARIATLPLFGILLIENVQRIINVRDVSPPVNAQMTVKDESLALPAVNPVY 1138  
 Qy 1211 TPQKSTLDNSLHKLQALGASGYTTSPPNINSVNAQSRGSLISTDSGNSLPERNS 1270

Db 1139 TPQKSTLDNLSLHKLGLGASIPYTTSTPINSVRNDRSGSLSTDSGNSLPERNS 1198  
QY 1271 EKSNSLDKQKQSSSTLGNVYRCCKLQSEKSLMCEFLYLKMSDMLFTYWNKASTSE 1330  
Db 1199 EKSNSLDKQKQSSSTLGNVYRCCKLQSEKSLMCEFLYLKMSDMLFTYWNKASTSE 1258  
QY 1331 LMDFTTISEVCLHQFOYMGKRYIA-----RTGMHARLQOLG 1367  
Db 1259 LMDFTTISEVCLHQFOYMGKRYIARNOEGIGFVHDKSGQTLFVSNRRGMMHARLQOLG 1318  
QY 1368 SLDNSLTFNHSHYSHSDADVLHOSLLEANLATEVCLTALDLSLFTLAFKQOLLADGHP 1427  
Db 1319 SLDNSLTFNHSHYSHSDADVLHOSLLEANLATEVCLTALDLSLFTLAFKQOLLADGHP 1378  
QY 1428 LMKVVDVYLCFLQKQSEETALKNVFTALRSLLYKFPSTPEGRADMCALCEYLKCN 1487  
Db 1379 LMKVVDVYLCFLQKQSEETALKNVFTALRSLLYKFPSTPEGRADMCALCEYLKCN 1438  
QY 1488 SKLSSIRTEASQLLYFLMRNFDYTKSKSFVTHLQVVISVQLADVVGIGTRFQSL 1547  
Db 1439 SKLSSIRTEASQLLYFLMRNFDYTKSKSFVTHLQVVISVQLADVVGIGTRFQSL 1498  
QY 1548 SIINNCANDRLIKHSPSSDVVDLTKRIRTVLMAATQKKEHNDPEMLVDLYSLAKSY 1607  
Db 1499 SIINNCANDRLIKHSPSSDVVDLTKRIRTVLMAATQKKEHNDPEMLVDLYSLAKSY 1558  
QY 1608 ASTPELRKTWLDSEMARHYKNGDLSAANCYHVTALVAEYLTKRAVOWEPPLPHSHS 1667  
Db 1559 ASTPELRKTWLDSEMARHYKNGDLSAANCYHVTALVAEYLTKR----- 1603  
QY 1668 ACLRRSGGVFRQCTAFRVTITENIDEBASMMEDVGMODVHFNEDVLMELLBOCADGLWK 1727  
Db 1604 -----GVFRQCTAFRVTITENIDEBASMMEDVGMODVHFNEDVLMELLBOCADGLWK 1655  
QY 1728 AERYELIADIYKLIITIEKRRPFRLAHLYDTLHAYSKVTBVMGSRLLGTIRVAF 1787  
Db 1656 AERYELIADIYKLIITIEKRRD----- 1678  
QY 1788 FGQAAQYQFTDSETEVEGFEDDGKEYIYKEPKLTPLEISORLLKLYSDKSGSENVNM 1847  
Db 1679 -----FFEDDGKEYIYKEPKLTPLEISORLLKLYSDKSGSENVNM 1720  
QY 1848 IODSGKVPKDLSDSKAYIQTHTVIFPDEKLOEKTEBERSHNIIRFMFEMPTQTK 1907  
Db 1721 IODSGKVPKDLSDSKAYIQTHTVIFPDEKLOEKTEBERSHNIIRFMFEMPTQTK 1780  
QY 1908 RQGVSEOCCKRRITLTAHCFPVYKRIIPVYQHTDNLPIEVALIDEMSKVAELRQLS 1967  
Db 1781 RQGVSEOCCKRRITLTAHCFPVYKRIIPVYQHTDNLPIEVALIDEMSKVAELRQLS 1840  
QY 1968 SAEVDMIKQLKQSSVSQVONAGPLAYARAFDDTNTKRYPNKYKLKEVRFQVEAC 2027  
Db 1841 SAEVDMIKQLKQSSVSQVONAGPLAYARAFDDTNTKRYPNKYKLKEVRFQVEAC 1900  
QY 2028 GQALAVNERLIKEDOLEYOEEMKANYREMAKELSEIMHQICPLEKTSVLPNSLHIFNA 2087  
Db 1901 GQALAVNERLIKEDOLEYOEEMKANYREMAKELSEIMHQICPLEKTSVLPNSLHIFNA 1960  
QY 2088 ISGTPSTWVHGMTSSSSVV 2107  
Db 1961 ISGTPSTWVHGMTSSSSVV 1980

## RESULT 4

ID AAB40315 standard; Protein; 1572 AA.

AC AAB40315;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF79 polypeptide sequence SEQ ID NO:158.

KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
KW vulnery; antiparasitic; antiparkinsonian; nocotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;  
KW immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antidiabetic;  
KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antidiabetic; antidiabetic; coagulation;  
KW thrombosis; contraceptive.  
OS Homo sapiens.  
PN WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000MO-US08621.  
PF 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CUBA-) CUBAGEN CORP.  
XX Shinkets RA, Leach M;  
PI MPI: 2000-602362/57.  
DR N-PSDB; AAC74524.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
PS Claim 11; Page 506-510; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiparasitic; antiparkinsonian; nocotropic; neuroprotective;  
CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antidiabetic; antiparasitic; antiparkinsonian; antineuritic; antithyroid;  
CC antidiabetic; antiparasitic; antiparkinsonian; antineuritic; antithyroid;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antidiabetic disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 1572 AA;  
Query Match 74.7%; Score 8173; DB 21; Length 1572;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1571; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 536 ASGNLDKARFSAITRQNSNKLSDNDMLKLADPKPKEMKALPYLGNLDTITNVSSD 595  
Db 1 ASGNLDKARFSAITRQNSNKLSDNDMLKLADPKPKEMKALPYLGNLDTITNVSSD 60

QY	596	PNNVNSYIPITKQETCSKPIPIFEVBEFUPCIBKHOPITYITNNHXYPKYXKJQDSQ	655
Db	61	FPNNVNSYIPITKQETCSKPIPIFEVBEFUPCIBKHOPITYITNNHXYPKYXKJQDSQ	120
QY	656	KSPAKARNIACIEPFKQDEDEDSQPLKCIYGRPGGVPVTRSAFAVLHHNNPEFYDEIK	715
Db	121	KSPAKARNIACIEPFKQDEDEDSQPLKCIYGRPGGVPVTRSAFAVLHHNNPEFYDEIK	180
QY	716	IELPTQJHEKHLILTPFVSCDSSKSTKKRDVETQVGSWMLPLKDGRRVYTSQHI	775
Db	181	IELPTQJHEKHLILTPFVSCDSSKSTKKRDVETQVGSWMLPLKDGRRVYTSQHI	240
QY	776	PVSNALPBGYLGQYOEJLGMGRHYGPEIKKVDGSKPLKISTILSVYVYQDHLNFPYC	835
Db	241	PVSNALPBGYLGQYOEJLGMGRHYGPEIKKVDGSKPLKISTILSVYVYQDHLNFPYC	300
QY	836	OKTESGALQENELVKYLSLHAMGHWIAFLPTILNQPRVLTIRATOEEVAVNVPVI	895
Db	301	OKTESGALQENELVKYLSLHAMGHWIAFLPTILNQPRVLTIRATOEEVAVNVPVI	360
QY	896	IHVVAQCHEGLSHLSRYKVAYKABEYVASBYKTVHEILTKSMTTILKESADFLTSNK	955
Db	361	IHVVAQCHEGLSHLSRYKVAYKABEYVASBYKTVHEILTKSMTTILKESADFLTSNK	420
QY	956	LLKTSWFFPDVLKSMAGHLIENSVKYLIRNORPASPASHAVETVNNMLMHIITQKPRDN	1015
Db	421	LLKTSWFFPDVLKSMAGHLIENSVKYLIRNORPASPASHAVETVNNMLMHIITQKPRDN	480
QY	1016	PEASNAHSLAVFIKRCFTFMDRGFVKQNNYISCPAIPDPKTLPEYKKEPRLRVCONH	1075
Db	481	PEASNAHSLAVFIKRCFTFMDRGFVKQNNYISCPAIPDPKTLPEYKKEPRLRVCONH	540
QY	1076	EHYIPLNLPMPEFGKRIORQDLOQDYSLTDEFCRNHPLVGLLRVETALQEBREYALI	1135
Db	541	EHYIPLNLPMPEFGKRIORQDLOQDYSLTDEFCRNHPLVGLLRVETALQEBREYALI	600
QY	1136	AISVUKNLLIKGSPDRYASRSHQARITVLPFLGLIENVQIRNTRDVSPFPVNAQT	1195
Db	601	AISVUKNLLIKGSPDRYASRSHQARITVLPFLGLIENVQIRNTRDVSPFPVNAQT	660
QY	1196	VKDESLAPANPLVTPQKSGTLDNSLHKDLDLGAISGASPYTSTPWINSVRNADSRGS	1255
Db	661	VKDESLAPANPLVTPQKSGTLDNSLHKDLDLGAISGASPYTSTPWINSVRNADSRGS	720
QY	1256	LISTDSGNSLPERNSEKSNLDKHOQOSTLGNVVRCPKLOOSEIKSLMCFLYILKSM	1315
Db	721	LISTDSGNSLPERNSEKSNLDKHOQOSTLGNVVRCPKLOOSEIKSLMCFLYILKSM	780
QY	1316	DDALFTYNNKASTSELMDFPTTISEVCLHQFOYMGKRYIARTGMMHARLQOLGSLDNSITF	1375
Db	781	DDALFTYNNKASTSELMDFPTTISEVCLHQFOYMGKRYIARTGMMHARLQOLGSLDNSITF	840
QY	1376	NHSYGHSDADVLHOSLEBANITATEVCLALDPLTSLFTLAFNQQLADHGNPLMKYFVDV	1435
Db	841	NHSYGHSDADVLHOSLEBANITATEVCLALDPLTSLFTLAFNQQLADHGNPLMKYFVDV	900
QY	1436	YLCFLQKQOSELAKNVFTALRSILYKRPSTFYEBRADWCAALCYEILKCNKSLKSISRT	1495
Db	901	YLCFLQKQOSELAKNVFTALRSILYKRPSTFYEBRADWCAALCYEILKCNKSLKSISRT	960
QY	1496	EASQULYFLMRNPNPYTKGKSFVRTHLOVITISVSQULAEVVGIGGTRFQOGLSIINNCA	1555
Db	961	EASQULYFLMRNPNPYTKGKSFVRTHLOVITISVSQULAEVVGIGGTRFQOGLSIINNCA	1020
QY	1556	SDRLIKHNSPSSDVLDLTKRIITYVMAAPOMKEHNDENMLVDLOYSLAKYASTPELRK	1615
Db	1021	SDRLIKHNSPSSDVLDLTKRIITYVMAAPOMKEHNDENMLVDLOYSLAKYASTPELRK	1080
QY	1616	TWLDMAIRIHVNGDLSAAMCYVAVTALVAEYLTRKSAVOMEPRLPHSHSACIRSRG	1675
Db	1081	TWLDMAIRIHVNGDLSAAMCYVAVTALVAEYLTRKSAVOMEPRLPHSHSACIRSRG	1140
QY	1676	GVFRQSCIAFRVITPNIDEBASWMDVGMDVHFNEDVIMELLBOCADGLMKABERYELIA	1735

Db	1141	GVFQGGCTAFRVLTIPNIDDEBSMMEDVGMDVHFNEVYLHLEBQCDGLMKERYELIA	1200
Qy	1736	DIYKLIIPIYKRRDFERLAHLVDTLHRAVSKYEVWHSGRRLIGTYFRVAFGQAAQYQ	1795
Db	1201	DIYKLIIPIYKRRDFERLAHLVDTLHRAVSKYEVWHSGRRLIGTYFRVAFGQAAQYQ	1260
Qy	1796	FTDSETDVEGFEEDEDEGKEYIYKEPKLTPLEISQRLKLYSXKFGSENVMIODSGKN	1855
Db	1261	FTDSETDVEGFEEDEDEGKEYIYKEPKLTPLEISQRLKLYSXKFGSENVMIODSGKN	1320
Qy	1856	PKDIDSKAVYQVTHVIPPEPEKELOERKTEPEESHNIIRBFMPMPPTONGKRGVBEQ	1915
Db	1321	PKDIDSKAVYQVTHVIPPEPEKELOERKTEPEESHNIIRBFMPMPPTONGKRGVBEQ	1380
Qy	1916	CKRRITILTAIHCFPYVKKRIIPVMYQHTTDNLPLEIYADENSKVYABIRLOLCSASAEVMIK	1975
Db	1381	CKRRITILTAIHCFPYVKKRIIPVMYQHTTDNLPLEIYADENSKVYABIRLOLCSASAEVMIK	1440
Qy	1976	LQKLQGSVSVQVNAAGPLAVAPAFLDNTNRYKYNKVKLLKEYFRQFVEACGQALAVNE	2035
Db	1441	LQKLQGSVSVQVNAAGPLAVAPAFLDNTNRYKYNKVKLLKEYFRQFVEACGQALAVNE	1500
Qy	2036	RLIPEQLEVOEEMKANYREMAKELSRIMEHQICPLEKTSVLPNSLIHIFAISGTPST	2095
Db	1501	RLIPEQLEVOEEMKANYREMAKELSRIMEHQICPLEKTSVLPNSLIHIFAISGTPST	1560
Qy	2096	MVHGTTSSSSSVV 2107	
Db	1561	MVHGTTSSSSSVV 1572	
RESULT 5			
ABG61676			
ID	ABG61676	standard; Protein; 1534 AA.	
AC	ABG61676;		
XX			
DT	13-AUG-2002	(first entry)	
XX			
DE		Cadherin-like asymmetry protein (CLASP) isoform #5.	
XX			
KM		Human; autoimmune disease; haematopoietic disorder; DiGeorge syndrome; blood protein disorder; agammaglobulinaemia; dysagammaglobulinaemia; ataxia telangiectasia; common variable immunodeficiency; lymphopenia; thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease; haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus; endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity; autoimmune pulmonary inflammation; organ rejection; inflammation; CLASP.	
XX			
XX		Homo sapiens.	
OS			
XX	WO200231117-A2.		
XX			
PD	18-APR-2002.		
XX			
PF	15-OCT-2001; 2001WO-US32202.		
XX			
PR	13-OCT-2000; 2000US-0687837.		
XX			
PA	(ARBO-) ARBOR VITA CORP.		
PA	(GARM/) GARMAN J D.		
PA	(CAND/) CANDIA A F.		
XX			
PI	Lu PS;		
XX			
DR	WP1; 2002-416861/44.		
XX			
PT	N-PSDB; ABK84970.		
XX			
New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating an immune response, and for treating multiple sclerosis, rheumatoid arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,			

PT and sepsis -  
 XX Discloure; Figure 3b; 245pp; English.  
 CC The invention relates to an isolated polypeptide (I) comprising an amino  
 XX acid sequence that has 90 % sequence identity to one of the human  
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2c, 2e)  
 CC sequences (PS). (I) is useful for identifying a compound or agent that  
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 CC inhibiting a immune response in a subject. A pharmaceutical composition  
 CC comprising a nucleic acid encoding (I) or (II) is useful for preventing  
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 CC the autoimmune disease is caused or exacerbated by increased activity  
 CC of Th1 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or  
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2  
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or  
 CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, DiGeorge syndrome, lymphopenia, thrombocytopenia, or  
 CC detect autoimmune diseases, e.g., Addison's disease, hemolytic anaemia,  
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,  
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary  
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or  
 CC hypersensitivity to an antigenic molecule, organ rejection or graft-  
 CC versus-host disease (GVHD) and inflammation. Abg61670-ABg61708  
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the  
 CC invention.  
 CC  
 XX  
 XX  
 SQ Sequence 1534 AA;  
 Query Match 72.9%; Score 7976; DB 23; Length 1534;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 301 QKTESGAQALGNELVYLKSLHAMEGHVMIAPLPTLNLQFRVLTRATQEEVAVNTRVI 360  
 QY IHVVAQCHBEGLSHLSRYVYKAYKAPYVASEYTVHEELTKSMTTLTKPSADLTSENK 955  
 DB 896 IHVVAQCHBEGLSHLSRYVYKAYKAPYVASEYTVHEELTKSMTTLTKPSADLTSENK 420  
 QY 956 LTKSWFFFDVLTKSMAOHLIENSVKLLRNQFPASVYHAAVEYVNMIMPHITOKFRDN 1015  
 DB 421 LTKSWFFFDVLTKSMAOHLIENSVKLLRNQFPASVYHAAVEYVNMIMPHITOKFRDN 480  
 QY 1016 PASANNAHSLAVFKRCPTFMDGFPVKQINNYISCAPGDPKTLFEYKEFFLAVYCNH 1075  
 DB 481 PASANNAHSLAVFKRCPTFMDGFPVKQINNYISCAPGDPKTLFEYKEFFLAVYCNH 540  
 QY 1076 EHYTFLNLPMPGKRIRYODLODYSLTDFECNHLVGLLREVGTLAOLDFEVRVLI 1135  
 DB 541 EHYTFLNLPMPGKRIRYODLODYSLTDFECNHLVGLLREVGTLAOLDFEVRVLI 600  
 QY 1136 AIVSLKNLLIKHSPDRYASRSHQARIATLYLPLFGLLIENVQINVDVSPFNAGMT 1195  
 DB 601 AIVSLKNLLIKHSPDRYASRSHQARIATLYLPLFGLLIENVQINVDVSPFNAGMT 660  
 QY 1196 VQDESLAPAVNPLVTPQKSTLDSNLKDLGALISGASPTTSTPINSVRNADSGS 1255  
 DB 661 VQDESLAPAVNPLVTPQKSTLDSNLKDLGALISGASPTTSTPINSVRNADSGS 720  
 QY 1256 LISTSGNSLPERNSEKNSLDKHOQSTLGNVYRCDKLDQSEIKSLMCFLYLTKSMS 1315  
 DB 721 LISTSGNSLPERNSEKNSLDKHOQSTLGNVYRCDKLDQSEIKSLMCFLYLTKSMS 780  
 QY 1316 DDALFTYNNKASTSELMDFITISEVCLHQFOYMGKRYIARTGMHARLQOLGSLDNLSTF 1375  
 DB 781 DDALFTYNNKASTSELMDFITISEVCLHQFOYMGKRYIARTGMHARLQOLGSLDNLSTF 840  
 QY 1376 NMSYGHSDADVLHOSLEENIATEVCLTALDTLSPTLAFKQULADGHNPLMKVVDV 1435  
 DB 841 NMSYGHSDADVLHOSLEENIATEVCLTALDTLSPTLAFKQULADGHNPLMKVVDV 900  
 QY 1436 YICFLQKQOSEFALKNVFTALRSLLYKPPSTFEGRADMCALCEYILKCKNSKLSIRT 1495  
 DB 901 YICFLQKQOSEFALKNVFTALRSLLYKPPSTFEGRADMCALCEYILKCKNSKLSIRT 960  
 QY 1496 EASQULYFLMRNPFDTGSKSFVRTHLQVLIISVSLADVVGIGTRFQOGLSIINNCAN 1555  
 DB 961 EASQULYFLMRNPFDTGSKSFVRTHLQVLIISVSLADVVGIGTRFQOGLSIINNCAN 1020  
 QY 1556 SPRLIKHTSFSSDVVDLTKRIRTVLMATQKKEHNDDEMLVDLOYSLAKSYASTPELRK 1615  
 DB 1021 SPRLIKHTSFSSDVVDLTKRIRTVLMATQKKEHNDDEMLVDLOYSLAKSYASTPELRK 1080  
 QY 1616 TWLDSMARIHVKNGLSEAMACYVHTALVAEYLTRKCAVQWEPPLPHSHSACLRSRG 1675  
 DB 1081 TWLDSMARIHVKNGLSEAMACYVHTALVAEYLTRKCAVQWEPPLPHSHSACLRSRG 1140  
 QY 1676 GVFRQGTAFRVITBNIDEASMMEDVGMQDVHFNEDVLMELLEQCADGLMAERYELIA 1735  
 DB 1141 GVFRQGTAFRVITBNIDEASMMEDVGMQDVHFNEDVLMELLEQCADGLMAERYELIA 1200  
 QY 1736 DIYKLIITIEYKRRPFERLAHLYDTLRAYSKYTEVWMSGRLLCTYRVAFFQGAAYO 1795  
 DB 1201 DIYKLIITIEYKRRPFERLAHLYDTLRAYSKYTEVWMSGRLLCTYRVAFFQGAAYO 1260  
 QY 1796 FTDSSETVEGFEDDGEKYEYKEPKLPLPSISORLLKLYSDKSGSENVKMIQDSGKN 1855  
 DB 1261 FTDSSETVEGFEDDGEKYEYKEPKLPLPSISORLLKLYSDKSGSENVKMIQDSGKN 1320  
 QY 1856 PKDLSKAYAYIQVTHVIPFDEKELQERKTEBERSHNIIRFMFENPFTQGRQGVBEQ 1915  
 DB 1321 PKDLSKAYAYIQVTHVIPFDEKELQERKTEBERSHNIIRFMFENPFTQGRQGVBEQ 1380  
 QY 1916 CKRRTILTAIHCFPVYKRIIPVMOHTDNLPIEVAIDEMSKVAELQLCSSAEDVMIK 1975  
 DB 1381 CKRRTILTAIHCFPVYKRIIPVMOHTDNLPIEVAIDEMSKVAELQLCSSAEDVMIK 1440

QY 1976 1QLKUGSVSVGNAGPLAVYARAFLDINTKRYPDNKKVLLKRVFQFYACGQALAVNE 2035  
 DB 1441 1QLKUGSVSVGNAGPLAVYARAFLDINTKRYPDNKKVLLKRVFQFYACGQALAVNE 1500  
 QY 2036 RLKEDQLEYQEMKANYEMAKELSEIMHEOI 2068  
 DB 1501 RLKEDQLEYQEMKANYEMAKELSEIMHEOI 1533

RESULT 6  
 ABG61680  
 ID ABG61680 standard. Protein, 1534 AA.  
 AC ABG61680;  
 DT 13-AUG-2002 (first entry)  
 DE Cadherin-like asymmetry protein (CLASP) isoform #9.  
 KW Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;  
 KW blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;  
 KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;  
 KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;  
 KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;  
 KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;  
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;  
 KW CLASP.  
 OS Homo sapiens.  
 PN WO20023117-A2.  
 XX 18-APR-2002.  
 PD 15-OCT-2001; 2001WO-US32202.  
 PF 13-OCT-2000; 2000US-0687837.  
 PR (ARBO-) ARBOR VITAE CORP.  
 PA (GARM/) GARMAN J D.  
 PA (CAND/) CANDIA A F.  
 XX Lu PI;  
 PI WPI: 2002-416861/44.  
 DR New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating  
 XX an immune response, and for treating multiple sclerosis, rheumatoid  
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,  
 PT and sepsis  
 CC -  
 XX Disclosure; Figure 5A; 245BP; English.  
 PS The invention relates to an isolated polypeptide (I) comprising an amino  
 XX acid sequence that has 90 % sequence identity to one of the human  
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2D)  
 CC sequences (PS). (I) is useful for identifying a compound or agent that  
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 CC inhibiting a immune response in a subject. A pharmaceutical composition  
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing  
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 CC the autoimmune disease is caused or exacerbated by increased activity  
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or  
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2  
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2

CC polynucleotides can screen for CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or  
 CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or  
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or  
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,  
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,  
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary  
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or  
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-  
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708  
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the  
 CC invention.

Sequence 1534 AA;  
 Query Match 72.9%; Score 7976; DB 23; Length 1534;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 536 ASGNLDKARPSAIVRQDSNKLSDNDMLKLADFPKPKMAKLPYLGNLDTITNVSSD 595  
 DB 1 ASGNLDKARPSAIVRQDSNKLSDNDMLKLADFPKPKMAKLPYLGNLDTITNVSSD 60  
 QY 596 PPNVNSSYIPDKOETCSKPTITFEVEEPCIPKHOPTIYNHLYVVPKYLYQSO 655  
 DB 61 PPNVNSSYIPDKOETCSKPTITFEVEEPCIPKHOPTIYNHLYVVPKYLYQSO 120  
 QY 656 KSFARARNAICIEFKDSDEBDSQPLKCIYGRPGSPVTRSAFAVLAHHQNPBYDRIK 715  
 DB 121 KSFARARNAICIEFKDSDEBDSQPLKCIYGRPGSPVTRSAFAVLAHHQNPBYDRIK 180  
 QY 716 IELPQLHEKHHLLTFPHVSGDNSSKSGTKKRDVETQVGYSMPLKDKGRVVTSEOH 775  
 DB 181 IELPQLHEKHHLLTFPHVSGDNSSKSGTKKRDVETQVGYSMPLKDKGRVVTSEOH 240  
 QY 776 PVSANLPBGYGYOELGKGRHYGPEIKWVDGKPLKISTHLVSVTYQDQHLNFPQYC 835  
 DB 241 PVSANLPBGYGYOELGKGRHYGPEIKWVDGKPLKISTHLVSVTYQDQHLNFPQYC 300  
 QY 836 QKTBSGAQALGNELVKYLSLHAMEGHVMIAPLPTILNLFRVLTFRATQEBVAVNTRVI 895  
 DB 301 QKTBSGAQALGNELVKYLSLHAMEGHVMIAPLPTILNLFRVLTFRATQEBVAVNTRVI 360  
 QY 896 IHVVAQCHEBGLSHLSRYKYAYKAPYVASEYKTVHEBITKSMITLIKDSADFLTSNK 955  
 DB 361 IHVVAQCHEBGLSHLSRYKYAYKAPYVASEYKTVHEBITKSMITLIKDSADFLTSNK 420  
 QY 956 LKTSWFFPDVLISMAOHLIENSVKVLRNORPASVHHAVETVNNMLPMPTITOKFPDN 1015  
 DB 421 LKTSWFFPDVLISMAOHLIENSVKVLRNORPASVHHAVETVNNMLPMPTITOKFPDN 480  
 QY 1016 PEASKNANSHLAVFIKRCFTFMDRGFVFKQNNYISCPAPDPKTLFEYKEEFLAVNCN 1075  
 DB 481 PEASKNANSHLAVFIKRCFTFMDRGFVFKQNNYISCPAPDPKTLFEYKEEFLAVNCN 540  
 QY 1076 EHYIPLNIMPFGKRIORYODLDVSLTDFCRNHLVGLLREVGTAQOEFREVRVI 1135  
 DB 541 EHYIPLNIMPFGKRIORYODLDVSLTDFCRNHLVGLLREVGTAQOEFREVRVI 600  
 QY 1136 AIVSVKLLIHGSPDRARASHOARIATLPLFGILLIENVORI NVADVSPFPNAGMT 1195  
 DB 601 AIVSVKLLIHGSPDRARASHOARIATLPLFGILLIENVORI NVADVSPFPNAGMT 660  
 QY 1196 VKDESLALPAVNPVTPQKGSTLDNSLHKDLGAIISGIASPYTTSTPNINSVRNADSRGS 1255

```

Db      661 VDESEALPAVPLVLPQKSGTSDNSLHDLGALIGIASPYTTSPTNINSVNMDSRS 720
      1256 LISTDSGNSLPERNSEKSNLDRQOOSTGNSVNCDLDOSEISLMLCFYLTKSMS 1315
      721 LISTDSGNSLPERNSEKSNLDRQOOSTGNSVNCDLDOSEISLMLCFYLTKSMS 780
      1316 DDALFTYMKASTSELMDFETISEVCLHOFQYMKRYIARTGMHARLOOLGSLDNLTF 1375
      781 DDALFTYMKASTSELMDFETISEVCLHOFQYMKRYIARTGMHARLOOLGSLDNLTF 840
      1376 NNSYGSDDADVHLQSLLENIAIEVCLTALDTLSLTLPKQOLADHGNPLMKVFPV 1435
      841 NNSYGSDDADVHLQSLLENIAIEVCLTALDTLSLTLPKQOLADHGNPLMKVFPV 900
      1436 YLCFLQKHOSETALKVFTALRSLLIKPSTFEGRADCAALCYEILKCNCKLSIRT 1495
      901 YLCFLQKHOSETALKVFTALRSLLIKPSTFEGRADCAALCYEILKCNCKLSIRT 960
      1496 EASQLLYPLMRNNDYTGKKSFVRTHLOVYISVSQIADVVGIGTRFOOSLSIINNCAN 1555
      961 EASQLLYPLMRNNDYTGKKSFVRTHLOVYISVSQIADVVGIGTRFOOSLSIINNCAN 1020
      1556 SDRLLKHTSESSSVKULTRIRTVLMTQMKHEHNDPMLVDLOSLAKSYASTPELRK 1615
      1021 SDRLLKHTSESSSVKULTRIRTVLMTQMKHEHNDPMLVDLOSLAKSYASTPELRK 1080
      1616 TMLDSMARLHVKNVGDSEAMCYVHTALVAEYLTRKEAVQMPPLPHSHSCLRRSRG 1675
      1081 TMLDSMARLHVKNVGDSEAMCYVHTALVAEYLTRKEAVQMPPLPHSHSCLRRSRG 1140
      1676 GVFRQCTAFRVITPNIIDEBASNMEDVGMQDVHFNEDVIMLEEQADGLWKAERYELIA 1735
      1141 GVFRQCTAFRVITPNIIDEBASNMEDVGMQDVHFNEDVIMLEEQADGLWKAERYELIA 1200
      1736 DIYKLIPIYEKRDDERLAHLVDTLHRAYSKTEVMHSGRLLGTYFRVAFPGQAAYQ 1795
      1201 DIYKLIPIYEKRDDERLAHLVDTLHRAYSKTEVMHSGRLLGTYFRVAFPGQAAYQ 1260
      1796 FTDSEIDVGEFFDEDEGKEYIYKEPLTPELSEISQRLKLYSDKFSSENVKMIQDSGKN 1855
      1261 FTDSEIDVGEFFDEDEGKEYIYKEPLTPELSEISQRLKLYSDKFSSENVKMIQDSGKN 1320
      1856 PKDLDSKAVYIQVTHVIPPEDEKELQERTKTEPERSHNIRPFMEPFTQGRKQGVBER 1915
      1321 PKDLDSKAVYIQVTHVIPPEDEKELQERTKTEPERSHNIRPFMEPFTQGRKQGVBER 1380
      1916 CKRRTILTALHCPYVKKRIPVWYQHNTDNLPEVAIDEMSKVVALRQLCSSAEVDMTK 1975
      1381 CKRRTILTALHCPYVKKRIPVWYQHNTDNLPEVAIDEMSKVVALRQLCSSAEVDMTK 1440
      1976 LQKLQGSVSVQVNAFLAYARAFLDPTTKRYPDKVKLLKEVFQFVACQALAVNE 2035
      1441 LQKLQGSVSVQVNAFLAYARAFLDPTTKRYPDKVKLLKEVFQFVACQALAVNE 1500
      2036 RLKEDQLEYOEEMKANYREMAKELSEIMHOL 2068
      1501 RLKEDQLEYOEEMKANYREMAKELSEIMHOL 1533

```

```

KW      chombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KW      haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
KW      endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
KW      autoimmune pulmonary inflammation; organ rejection; inflammation;
KW      CLASP.
      OS      Homo sapiens.
      PN      WO200231117-A2.
      PD      18-APR-2002.
      PF      15-OCT-2001; 2001WO-US32202.
      PR      13-OCT-2000; 2000US-0687837.
      PA      (ARBO-) ARBOR VITA CORP.
      PA      (GARM/) GARMAN J D.
      PA      (CAND/) CANDIA A F.
      PI      Lu PS;
      PI      WPI; 2002-416861/44.
      DR      New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
      PT      an immune response, and for treating multiple sclerosis, rheumatoid
      PT      arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
      PT      and sepsis
      PS      Example 4; Figure 6B; 245pp; English.
      XX      The invention relates to an isolated polypeptide (I) comprising an amino
      CC      acid sequence that has 90 % sequence identity to one of the human
      CC      cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
      CC      sequences (PS). (I) is useful for identifying a compound or agent that
      CC      binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
      CC      detecting a CLASP-2 polypeptide in a sample. (II) is useful for
      CC      inhibiting a immune response in a subject. A pharmaceutical composition
      CC      comprising a nucleic acid encoding (I), or (II) is useful for preventing
      CC      or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
      CC      the autoimmune disease is caused or exacerbated by increased activity
      CC      of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
      CC      detection or inhibition of CLASP-2 expression (e.g., antisense or
      CC      ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
      CC      polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
      CC      antibodies or are used as therapeutic polypeptides. The CLASP-2
      CC      polynucleotide or fragments can be used in diagnostics (e.g., as probes
      CC      for CLASP-2 expression), as a lymphocyte marker and for therapeutic
      CC      purposes. CLASP-2 polynucleotides can construct transgenic and knockout
      CC      animals, e.g., for screening for CLASP-2 agonists and antagonists. CLASP-2
      CC      polynucleotides can screen for CLASP-2 agonists and antagonists.
      CC      CLASP-2 polypeptides or polynucleotides can treat deficiencies or
      CC      disorders of the immune system, by activating or inhibiting the
      CC      activation, differentiation of immune cells and can treat or detect
      CC      deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
      CC      or polynucleotides can increase differentiation and proliferation of
      CC      haematopoietic cells, including the pluripotent stem cells to treat those
      CC      disorders associated with a decrease in certain (or many) types of
      CC      haematopoietic cells e.g., immunologic deficiency syndromes including
      CC      blood protein disorders (e.g., agammaglobulinemia,
      CC      dysgammaglobulinemia, ataxia telangiectasia, common variable
      CC      immunodeficiency, DiGeorge syndrome, lymphopenia, thombocytopenia, or
      CC      haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
      CC      detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
      CC      Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
      CC      endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
      CC      inflammation. CLASP-2 can be used to treat anaphylaxis or
      CC      hypersensitivity to an antigenic molecules, organ rejection or graft-
      CC      versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
      CC      represent cadherin-like asymmetry protein (CLASP) sequences of the
      CC      invention.
      XX      Sequence 1534 AA;
      SQ

```

Query Match	72.9%; Score 7976; DB 23; Length 1534;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1532; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
QY	536 ASGNLDKARFSAIIRKQDSNKLSDNDMLKLLADFKREBKAKLPVILGMLDITIDNVSSD 595
DB	1 ASGNLDKARFSAIIRKQDSNKLSDNDMLKLLADFKREBKAKLPVILGMLDITIDNVSSD 60
QY	596 PNNVNVSSYIPTKOPETCKPTTFEVEBFVPCIPGHQPYITTYTNHLVYKYLKYDSO 655
DB	61 PNNVNVSSYIPTKOPETCKPTTFEVEBFVPCIPGHQPYITTYTNHLVYKYLKYDSO 120
QY	656 KSPAKARNIAICIEFKDSDEDSQPLKCIYGRPGCVFTTRSAFAVALHNNQDEFEYDEIK 715
DB	121 KSPAKARNIAICIEFKDSDEDSQPLKCIYGRPGCVFTTRSAFAVALHNNQDEFEYDEIK 180
QY	716 IELPTOLHEKHHLLTFPHVSCDSSKSGSTKKRDVETQYGYSMPLKDGGRVVTSEQH 775
DB	181 IELPTOLHEKHHLLTFPHVSCDSSKSGSTKKRDVETQYGYSMPLKDGGRVVTSEQH 240
QY	776 PVSANLPSGVLGYOELGMRHYPKIKWVDGKPKLKISTHUVSTVYTDQHLNFFOYC 835
DB	241 PVSANLPSGVLGYOELGMRHYPKIKWVDGKPKLKISTHUVSTVYTDQHLNFFOYC 300
QY	836 QKTESGAQALGNELVYKLSLHAMEGHWIAFLPTILNQLFRVLTATOEVAVAVNTRYI 895
DB	301 QKTESGAQALGNELVYKLSLHAMEGHWIAFLPTILNQLFRVLTATOEVAVAVNTRYI 360
QY	896 IHVAVACHEEGLESILRSYKCAKYKAPYVASEYKVEHELTSMWTTILKPSADFLTSNK 955
DB	361 IHVAVACHEEGLESILRSYKCAKYKAPYVASEYKVEHELTSMWTTILKPSADFLTSNK 420
QY	956 LKXWMEFFEDVLIKMAOHLIENSKYKILRNORFPAHYHAEVTVNMMLPHITQKERPN 1015
DB	421 LKXWMEFFEDVLIKMAOHLIENSKYKILRNORFPAHYHAEVTVNMMLPHITQKERPN 480
QY	1016 PEASIKANHSIAVFIKRCCTFMDRGFVFKQINNYISCPAPGPKTLFEYKFEPLRVACNH 1075
DB	481 PEASIKANHSIAVFIKRCCTFMDRGFVFKQINNYISCPAPGPKTLFEYKFEPLRVACNH 540
QY	1076 EHYIPLNLPMPFGKRIQRYYQDLQDLSLTDEFGRNHPVGLLLREVGTALEQEFREVL 1135
DB	541 EHYIPLNLPMPFGKRIQRYYQDLQDLSLTDEFGRNHPVGLLLREVGTALEQEFREVL 600
QY	1136 AISVLKNLLIKHSFDRVARSRSHQARIATLYPLFGLLENNQRIINRVSPPPVNAAGT 1195
DB	601 AISVLKNLLIKHSFDRVARSRSHQARIATLYPLFGLLENNQRIINRVSPPPVNAAGT 660
QY	1196 VDESLALPAAVNPVLPQKGSTLDNGLKDLGAIASPIYTTSTPININSYRNADSRGS 1255
DB	661 VDESLALPAAVNPVLPQKGSTLDNGLKDLGAIASPIYTTSTPININSYRNADSRGS 720
QY	1256 LISTDSGNSLPERNSEKNSLDKQGSSTLGNSVACDLKDOSEIKSLMCELYIILKNS 1315
DB	721 LISTDSGNSLPERNSEKNSLDKQGSSTLGNSVACDLKDOSEIKSLMCELYIILKNS 780
QY	1316 DDALFTYNNKASTSEIMDFETISEVCLHQPYWNGKRYIARTGMHARLQOLGSLDNLSTF 1375
DB	781 DDALFTYNNKASTSEIMDFETISEVCLHQPYWNGKRYIARTGMHARLQOLGSLDNLSTF 840
QY	1376 NNSYGHSDADVLAHQSLLEANIATEVCLTALDLSLTFLAFKQQLADHGHNPMLKKVAFDV 1435
DB	841 NNSYGHSDADVLAHQSLLEANIATEVCLTALDLSLTFLAFKQQLADHGHNPMLKKVAFDV 900
QY	1436 YICFLQKHQSEETALKVFTALBSLIYKPSSTFYEGALMCAALCYEILKCCNSKLSIRT 1495
DB	901 YICFLQKHQSEETALKVFTALBSLIYKPSSTFYEGALMCAALCYEILKCCNSKLSIRT 960
QY	1496 EASQLLYFLMRNPFDTGKKSFPRTHLQVYISVQLIADVNGIGTRFOQSLSIINNCA 1555
DB	961 EASQLLYFLMRNPFDTGKKSFPRTHLQVYISVQLIADVNGIGTRFOQSLSIINNCA 1020

QY	1556 SDRLIKHTSPSSDVADLTIRKIRITVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRK 1615
DB	1021 SDRLIKHTSPSSDVADLTIRKIRITVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRK 1080
QY	1616 TWLDSMAIITHVNGSLSEAAKCYAVNTALVAEYLTRKAAVQMBPPLPHSHACLRSSG 1675
DB	1081 TWLDSMAIITHVNGSLSEAAKCYAVNTALVAEYLTRKAAVQMBPPLPHSHACLRSSG 1140
QY	1676 GVFRQCTAFRITNTNIBEASMMEDVGMQDVHFNEVDYLMELLBOCADGLMAERYELLA 1735
DB	1141 GVFRQCTAFRITNTNIBEASMMEDVGMQDVHFNEVDYLMELLBOCADGLMAERYELLA 1200
QY	1736 DIYKLIIPYERKRDPERLAHLDTLHRAYSKTEVMSGRLLGTFRRVAFPGQAAQYQ 1795
DB	1201 DIYKLIIPYERKRDPERLAHLDTLHRAYSKTEVMSGRLLGTFRRVAFPGQAAQYQ 1260
QY	1796 FTDSSTVDEGFEDEDGKEYIYKEKLTPLSEISQRLKLYSDKGSBNVKKIIDSQKN 1855
DB	1261 FTDSSTVDEGFEDEDGKEYIYKEKLTPLSEISQRLKLYSDKGSBNVKKIIDSQKN 1320
QY	1856 PKDLSKTAIYQVTHVIPFPDEKELQERKTEPERSHNIRPFEMPFTOTGRQGSVEBQ 1915
DB	1321 PKDLSKTAIYQVTHVIPFPDEKELQERKTEPERSHNIRPFEMPFTOTGRQGSVEBQ 1380
QY	1916 CKRRITLTAIHCFPYVKRIPVWYQHTDNLPIEVALIDEMSKVAELRQLCSSAEVDIK 1975
DB	1381 CKRRITLTAIHCFPYVKRIPVWYQHTDNLPIEVALIDEMSKVAELRQLCSSAEVDIK 1440
QY	1976 LQKIQGSYVQVNAAGPLAFAFLDDNTTKYIPNNKYKLKEVROPYACGALVANE 2035
DB	1441 LQKIQGSYVQVNAAGPLAFAFLDDNTTKYIPNNKYKLKEVROPYACGALVANE 1500
QY	2036 RLKEDOLEYOEMKANYREMAKELSEIMHEQI 2068
DB	1501 RLKEDOLEYOEMKANYREMAKELSEIMHEQI 1533
RESULT 8	
AAB36527	
ID	AAB36527 standard; Protein; 1352 AA.
XX	
AC	AAB36527;
XX	
DT	07-MAR-2001 (first entry)
XX	
DE	Human CLASP-2 protein sequence.
XX	
KW	CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory;
KW	cadherin-like asymmetry protein; autoimmune disease; immunosuppressive;
KW	immunomodulatory; antiinflammatory; antirheumatic; cyclostatic;
KW	hypertensive; antirheumatic; antianemic; haemostatic; neuroprotective;
KW	hypersensitivity; transplantation rejection response; immunodeficiency;
KW	proliferation; differentiation; inflammatory response; arthritis;
KW	inflammatory bowel disease; haematopoietic cell; blood protein disorder;
KW	anaemia; thrombocytopaenia; multiple sclerosis; rheumatoid arthritis;
KW	endometriosis; pregnancy induced hypertension.
OS	Homo sapiens.
XX	
PN	WO200061747-A2.
XX	
PD	19-OCT-2000.
XX	
PF	13-APR-2000; 2000MO-US10158.
XX	
PR	14-APR-1999; 99US-0129171.
XX	
PR	14-MAY-1999; 99US-0134114.
XX	
PR	14-MAY-1999; 99US-0134117.
XX	
PR	14-MAY-1999; 99US-0134118.
XX	
PR	21-OCT-1999; 99US-0160860.
XX	
PR	29-OCT-1999; 99US-0162498.
XX	
PR	13-DEC-1999; 99US-0170453.
XX	
PR	14-JAN-2000; 2000US-0176195.







07-MAR-2001 (first entry)  
Human CLASP-2A protein sequence.  
CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory; cadherin-like asymmetry protein; autoimmune disease; immunosuppressive; immunomodulatory; antiinflammatory; antileukemic; cytostatic; hypotensive; antirheumatic; antineoplastic; haemostatic; neuroprotective; hyperproliferative; transplantation rejection response; immunodeficiency; proliferation; differentiation; inflammatory response; arthritis; inflammatory bowel disease; haematopoietic cell; blood protein disorder; anaemia; thrombocytopenia; multiple sclerosis; rheumatoid arthritis; endometriosis; pregnancy induced hypertension.  
Homo sapiens.  
WO200061747-A2.  
19-OCT-2000.  
13-APR-2000; 2000WO-US10158.  
14-APR-1999; 99US-0129171.  
14-MAY-1999; 99US-0134114.  
14-MAY-1999; 99US-0134117.  
14-MAY-1999; 99US-0134118.  
21-OCT-1999; 99US-0160860.  
29-OCT-1999; 99US-0162498.  
13-DEC-1999; 99US-0170453.  
14-JAN-2000; 2000US-0176195.  
14-FEB-2000; 2000US-0182296.  
(ARBO-) ARBOR VITA CORP.  
Lu PS;  
WPI; 2000-619230/59.  
N-PSDB; AAC87973.  
Isolated cadherin-like asymmetry protein-2 polynucleotide and polypeptide used to diagnose, treat and prevent autoimmune diseases and inflammatory responses -  
Example 1; Fig 2B; 286bp; English.  
The present invention describes cadherin-like asymmetry protein-2 (CLASP-2). CLASP-2 can have immunosuppressive, immunomodulatory, antiinflammatory, antileukemic, cytostatic, hypotensive, antineoplastic, antineoplastic, haemostatic and neuroprotective activities. CLASP-2 can be used to inhibit an immune response in a subject by interfering with the ability of a CLASP-2 protein to bind to another T cell or B cell. An immune response in a subject may also be inhibited by administering an antibody which specifically binds to CLASP-2. CLASP-2 polynucleotides, proteins and antibodies can be used to prevent or treat a CLASP-2 mediated disease, such as an autoimmune disease caused or exacerbated by increased activity of TH1 cells. They can also be used to treat hypersensitivities, prevent transplantation rejection responses and augment immune responsiveness in immunodeficiency states, inhibit proliferation and differentiation of cells involved in an inflammatory response, e.g. arthritis, inflammatory bowel disease and increase differentiation and proliferation of haematopoietic cells e.g. to treat anaemia, thrombocytopenia and other blood protein disorders. Disorders treated by disrupting CLASP-2 function include multiple sclerosis, rheumatoid arthritis, endometriosis and pregnancy induced hypertension. The present sequence represents human CLASP-2A, which is used in the exemplification of the present invention.

Query Match 62.0%; Score 6784.5; DB 21; Length 1352;  
Best Local Similarity 92.7%; Pred. No. 0;  
Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLHHHNPDEYDEIKIELPTQLHEKHLLTFPHVSCNSSKSGTKRQDVETQVGSWL 760  
Db 1 VLHHHNPDEYDEIKIELPTQLHEKHLLTFPHVSCNSSKSGTKRQDVETQVGSWL 60  
QY 761 PLKXGRVVTSEOHIPVANI PSGTGYQELGMRHVPETIKMDGGRPLKISTHVLST 820  
Db 61 PLKXGRVVTSEOHIPVANI PSGTGYQELGMRHVPETIKMDGGRPLKISTHVLST 120  
QY 821 VYTDQHLNHFQYCKQKESGAGLNLVYKLSLHMEGHVMAFLPTLLNQLFRVLT 880  
Db 121 VYTDQHLNHFQYCKQKESGAGLNLVYKLSLHMEGHVMAFLPTLLNQLFRVLT 180  
QY 881 RATQEEVAVNTVRIIVHVAQCHEEGESLRSYKVAKYKAPVYASRYKTVHEELYSM 940  
Db 181 RATQEEVAVNTVRIIVHVAQCHEEGESLRSYKVAKYKAPVYASRYKTVHEELYSM 240  
QY 941 TTILKPSADPLTSNKLKYSWFFPDVLIKMAOHLIENSKYKTLRNGCFPPASYNHAAETV 1000  
Db 241 TTILKPSADPLTSNKLKYSWFFPDVLIKMAOHLIENSKYKTLRNGCFPPASYNHAAETV 300  
QY 1001 VNMLPHITTOKFRDNPESKANSLSLAVFIKRCFTFMDRGVFKOINNYISCFAPGDPRT 1060  
Db 301 VNMLPHITTOKFRDNPESKANSLSLAVFIKRCFTFMDRGVFKOINNYISCFAPGDPRT 360  
QY 1061 LFEYKFEFLRVYCNHEHYIPLNLPMPFGKRIQRYQDQLDYSLTDFECRNHFLVGLLLR 1120  
Db 361 LFEYKFEFLRVYCNHEHYIPLNLPMPFGKRIQRYQDQLDYSLTDFECRNHFLVGLLLR 420  
QY 1121 EVGTALQFREVRLAIVLKNLTKHSFDDRYASRSHQAATATYLPFLGLIENVORI 1180  
Db 421 EVGTALQFREVRLAIVLKNLTKHSFDDRYASRSHQAATATYLPFLGLIENVORI 480  
QY 1181 NVRDVSPPPVNAAGMTVKDESIALPAVNPLVTPQKSTLDNSLKHDLAIGIASPYTTS 1240  
Db 481 NVRDVSPPPVNAAGMTVKDESIALPAVNPLVTPQKSTLDNSLKHDLAIGIASPYTTS 540  
QY 1241 TENINSVANADSRGSLISTDSGNSLPERNSEKNSLDKHOQSGTIGNSVVRCDKLDQSEI 1300  
Db 541 TENINSVANADSRGSLISTDSGNSLPERNSEKNSLDKHOQSGTIGNSVVRCDKLDQSEI 600  
QY 1301 KSLMCFLLYILKNSDDALFTYWNKASTSELMDFITSEVCLHQPQYWGKRYIA----- 1354  
Db 601 KSLMCFLLYILKNSDDALFTYWNKASTSELMDFITSEVCLHQPQYWGKRYIA----- 660  
QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLEANIA 1397  
Db 661 GPVYHDKRSQTLPVSRNFTGMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLEANIA 720  
QY 1398 TEVCUTALDTLSLFTLAFKQNLADHGNPLMKVFDVYLCPLOKHOSETALKNVFTALR 1457  
Db 721 TEVCUTALDTLSLFTLAFKQNLADHGNPLMKVFDVYLCPLOKHOSETALKNVFTALR 780  
QY 1458 SLIYFPESTFYEGRADMCALCYEILKCCNSKLSIRTEASQQLYFLMRNPFDTYKGSF 1517  
Db 781 SLIYFPESTFYEGRADMCALCYEILKCCNSKLSIRTEASQQLYFLMRNPFDTYKGSF 840  
QY 1518 VRTHLQVITISQILADVVGIGTRFOOSLIINNCAUSDLLIKHTSSSDVKDLTKIR 1577  
Db 841 VRTHLQVITISQILADVVGIGTRFOOSLIINNCAUSDLLIKHTSSSDVKDLTKIR 900  
QY 1578 TVLWATAQMKHENDPEMLVDLOYSIAKSYASTBELRTKLWDSMARIHVKNGDLSAAMC 1637  
Db 901 TVLWATAQMKHENDPEMLVDLOYSIAKSYASTBELRTKLWDSMARIHVKNGDLSAAMC 960  
QY 1638 YVHTALVAEYLTREKAVQWEPPLPHSHSACLRRSGCVFRQGTARVITPINIDBAS 1697  
Db 961 YVHTALVAEYLTREKAVQWEPPLPHSHSACLRRSGCVFRQGTARVITPINIDBAS 997  
QY 1698 MMEDEVGMQDVHFNEDVLMELLECAGDGLMKABRYELADIYKLIIFIEYKRRDPERLNL 1757  
Db 998 MMEDEVGMQDVHFNEDVLMELLECAGDGLMKABRYELADIYKLIIFIEYKRRDP----- 1051  
QY 1758 YDTLHRAYSKTEVWHSGRLLGTYFRVAFPGOAAQOYFTDSETDVEGFPDEBDGKEYIY 1817

```

Db      1052 ----- FEDEGKEKYY 1062
Qy      1818 KEPKLPLSEISORLLKLYSDKFGSENVMMIODSGKVPKDLDSKXAYIQVTHVPEFPE 1877
Db      1063 KEKPLPLSEISORLLKLYSDKFGSENVMMIODSGKVPKDLDSKXAYIQVTHVPEFPE 1122
Qy      1878 KELOEKTEFERSHNRIRFMFEMPTQTGKRGQVEEOCKRRTILTAHCFFPKKIRIV 1937
Db      1123 KELOEKTEFERSHNRIRFMFEMPTQTGKRGQVEEOCKRRTILTAHCFFPKKIRIV 1182
Qy      1938 MYOHTDLPPIEVAIDEMSKKVAELQOLCSAEVDMIKIQKQGSVSVQVNAGPLAVAR 1997
Db      1183 MYOHTDLPPIEVAIDEMSKKVAELQOLCSAEVDMIKIQKQGSVSVQVNAGPLAVAR 1242
Qy      1998 AFLDDNTKRYPNKVKLLKEVROFVEACGALANERLIKEDOLEYOEEMKANVREMA 2057
Db      1243 AFLDDNTKRYPNKVKLLKEVROFVEACGALANERLIKEDOLEYOEEMKANVREMA 1302
Qy      2058 KEISEIMHBOICPLEEKTSVLPNSLHPNAISGTPSTVHWGHTSSSSV 2107
Db      1303 KEISEIMHBOICPLEEKTSVLPNSLHPNAISGTPSTVHWGHTSSSSV 1352

RESULT 10
ABG61670
ID ABG61670 standard; Protein; 1353 AA.
XX
AC ABG61670;
XX
DT 13-AUG-2002 (first entry)
XX
DE Cadherin-like asymmetry protein (CLASP).
XX
KW Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;
KW blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;
KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
KW autoimmune pulmonary inflammation; organ rejection; inflammation;
KW CLASP.
XX
OS Homo sapiens.
XX
PN MO20023117-A2.
XX
PD 18-APR-2002.
XX
PF 15-OCT-2001; 2001WO-US32202.
XX
PR 13-OCT-2000; 2000US-0687837.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PA (GARM/) GARMAN J D.
XX
PA (CAND/) CANDIA A F.
XX
PI Lu PS;
XX
DR WPI: 2002-416861/44.
DR N-PSDB: ABR84964.
XX
PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT an immune response, and for treating multiple sclerosis, rheumatoid
PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT and sepsis
XX
PS Disclosure; Figure 1; 245pp; English.
XX
CC The invention relates to an isolated polypeptide (I) comprising an amino
CC acid sequence that has 90 % sequence identity to one of the human
CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
CC sequences (PS). (I) is useful for identifying a compound or agent that

```

```

CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
CC inhibiting a immune response in a subject. A pharmaceutical composition
CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
CC the autoimmune disease is caused or exacerbated by increased activity
CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
CC detection or inhibition of CLASP-2 expression (e.g., antisense or
CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
CC antibodies or are used as therapeutic polypeptides. The CLASP-2
CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
CC polynucleotides can screen for CLASP-2 agonists and antagonists.
CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
CC disorders of the immune system, by activating or inhibiting the
CC activation, differentiation of immune cells and can treat or detect
CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
CC or polynucleotides can increase differentiation and proliferation of
CC haematopoietic cells, including the pluripotent stem cells to treat those
CC disorders associated with a decrease in certain (or many) types of
CC haematopoietic cells e.g., immunologic deficiency syndromes including
CC blood protein disorders (e.g., agammaglobulinemia,
CC dysgammaglobulinemia, ataxia telangiectasia, common variable
CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or
CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
CC inflammation. CLASP-2 can be used to treat anaphylaxis or
CC hypersensitivity to an antigenic molecules, organ rejection or graft-
CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61108
CC represent cadherin-like asymmetry protein (CLASP) sequences of the
CC invention.
XX
SQ Sequence 1353 AA;
XX
Query Match 62.0%; Score 6784.5; DB 23; Length 1353;
Beet Local Similarity 92.7%; Pred. No. 0;
Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;
Qy 701 VLHHNQNEPEYDEIKELPTQLHEKHLLTFHFVSCNNSKSGSKKRDVETQGVGYML 760
Db 1 VLHHNQNEPEYDEIKELPTQLHEKHLLTFHFVSCNNSKSGSKKRDVETQGVGYML 60
Qy 761 PLKDGRRVVTSEOHIPVANSIPSGYLGYOELGMRHYPETIKWDGKPLKISTHLYST 820
Db 61 PLKDGRRVVTSEOHIPVANSIPSGYLGYOELGMRHYPETIKWDGKPLKISTHLYST 120
Qy 821 VYTDOPHLNPFQYCKTESGAQALGNELVYKLSLHMBEHWMAFLPTLNOLFRLVT 880
Db 121 VYTDOPHLNPFQYCKTESGAQALGNELVYKLSLHMBEHWMAFLPTLNOLFRLVT 180
Qy 881 RATOEEVAVNTTRVLIHVAAQCHEGLESHLSRYKYVKAEPYVASEYKTHBELTSM 940
Db 181 RATOEEVAVNTTRVLIHVAAQCHEGLESHLSRYKYVKAEPYVASEYKTHBELTSM 240
Qy 941 TTILKPSADPLTSNKLKTSWPPFDVLTKSMAOHLIENSKYKLLRNQFPASVYHNAVETV 1000
Db 241 TTILKPSADPLTSNKLKTSWPPFDVLTKSMAOHLIENSKYKLLRNQFPASVYHNAVETV 300
Qy 1001 VNMMPHITQKFRPNPEASKNANSIIVIRCTFPMRGFVFOINNYISCPAGDPKT 1060
Db 301 VNMMPHITQKFRPNPEASKNANSIIVIRCTFPMRGFVFOINNYISCPAGDPKT 360
Qy 1061 LFEYKFEFLRVVCHNEHTIPLNMPFGKGRIRQYQDQLDYSLTDEFCRNHFLVGLLIR 1120
Db 361 LFEYKFEFLRVVCHNEHTIPLNMPFGKGRIRQYQDQLDYSLTDEFCRNHFLVGLLIR 420
Qy 1121 EVGTALQEFREVRLLAISVLKLLIKHSFDRVARSRHOARIATLYLPFGILLIENVORI 1180

```

421 EVGTALQEFREVLIAISVLKNLLIKHSFDRYVASHQARIATITVLPFGLLIENVORI 480  
 QY 1181 NVRDVSPFPVNAAGMTYKDESLAPAVNPVLTPOKSTLDNSLHKDLGASIGIAPYTT 1240  
 DB 481 NVRDVSPFPVNAAGMTYKDESLAPAVNPVLTPOKSTLDNSLHKDLGASIGIAPYTT 540  
 QY 1241 TENINSVRNADSRGSLISTDSGNSLPERNSEKSNLIDKHCQSSITGNSVVRCDLQDSEI 1300  
 DB 541 TENINSVRNADSRGSLISTDSGNSLPERNSEKSNLIDKHCQSSITGNSVVRCDLQDSEI 600  
 QY 1301 KSLLMCFYIILKSMGDDALFTYWNKASTSELMDFPISYCLHOCQYMGKRIYA----- 1354  
 DB 601 KSLLMCFYIILKSMGDDALFTYWNKASTSELMDFPISYCLHOCQYMGKRIYA----- 1354  
 QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLEANIA 1397  
 DB 661 GPVYDRKSQTLPSGRNRTGMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLEANIA 1457  
 QY 1398 TEVCUTALDTLSLFTLAFKNQDLADHGNPLMKKVPDYVJLCFLQKHQSEALKNVFTALR 1457  
 DB 721 TEVCUTALDTLSLFTLAFKNQDLADHGNPLMKKVPDYVJLCFLQKHQSEALKNVFTALR 1457  
 QY 1458 SLIYFPSTFYEGRADMCALCYEILKCNKSLSTRTASQULFLMKNNDYTGKSP 1517  
 DB 781 SLIYFPSTFYEGRADMCALCYEILKCNKSLSTRTASQULFLMKNNDYTGKSP 1517  
 QY 1518 VATHQVLIISVSQLADVVGIGTRFOOSLTIINNCANSRLIKHTSPSSDYKDLTKRIR 1577  
 DB 841 VATHQVLIISVSQLADVVGIGTRFOOSLTIINNCANSRLIKHTSPSSDYKDLTKRIR 1577  
 QY 1578 TVLMAQAQKHEHNDPEMLVDLOYSLAKSVASTPELRKTLMDSMARHVKNDLSEAMC 1637  
 DB 901 TVLMAQAQKHEHNDPEMLVDLOYSLAKSVASTPELRKTLMDSMARHVKNDLSEAMC 1637  
 QY 1638 YHATVALVAEYITRKBAVOMEPRLPHSHSACLRSSRGCVFPOGCTAFVITPNIIDEAS 1697  
 DB 961 YHATVALVAEYITRKBAVOMEPRLPHSHSACLRSSRGCVFPOGCTAFVITPNIIDEAS 1697  
 QY 1698 NMEDVGMODVHFNEDVLMELBOCADGLMKARVELIADYLIPIYKRRDFERLAML 1757  
 DB 998 NMEDVGMODVHFNEDVLMELBOCADGLMKARVELIADYLIPIYKRRDFERLAML 1757  
 QY 1758 YDTLHRAVSKTVHMSGRRLGTYFRVAFGQAQYQITDSEDTVEGFEDEDEGKEYIY 1817  
 DB 1052 -----FEDEDEGKEYIY 1817  
 QY 1818 KEPKLTPLUSEISQRLIKLYSKRGSENVMIODSGKNPKDLSKYAYIQTTHVIFPDE 1877  
 DB 1063 KEPKLTPLUSEISQRLIKLYSKRGSENVMIODSGKNPKDLSKYAYIQTTHVIFPDE 1877  
 QY 1878 KELOARKTEFEHSHNIRRMPEMPPTQNGKRGVBEQCKRRTILTAHCFPVYKRRIPV 1937  
 DB 1123 KELOARKTEFEHSHNIRRMPEMPPTQNGKRGVBEQCKRRTILTAHCFPVYKRRIPV 1937  
 QY 1938 MYQHTDLNPIEVAIDEMSKVAELRQLCSSAEVMIKQLKQSSVSQVWAGPLAYAR 1997  
 DB 1183 MYQHTDLNPIEVAIDEMSKVAELRQLCSSAEVMIKQLKQSSVSQVWAGPLAYAR 1997  
 QY 1998 AFLDOTNTKRYPDNKKYKLYKEVPROFEACGALLAVNEFLIYEDOLEYOEMKANYREMA 2057  
 DB 1243 AFLDOTNTKRYPDNKKYKLYKEVPROFEACGALLAVNEFLIYEDOLEYOEMKANYREMA 2057  
 QY 2058 KELSIMEHOICPLEEKTSLVNSLIHFAISGTPSTVWGHMTSSSVV 2107  
 DB 1303 KELSIMEHOICPLEEKTSLVNSLIHFAISGTPSTVWGHMTSSSVV 2107

RESULT 11  
 ABG61671 ID ABG61671 standard; Protein; 1353 AA.  
 AC ABG61671;  
 XX

13-AUG-2002 (first entry)  
 XX  
 DE Cadherin-like asymmetry protein (CLASP) isoform.  
 XX  
 KM Human; autoimmune disease; haematopoietic disorder; Disgeorge syndrome;  
 KM blood protein disorder; agammaglobulinemia; dysgammaaglobulinemia;  
 KM ataxia telangiectasia; common variable immunodeficiency; lymphopenia;  
 KM thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;  
 KM haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;  
 KM endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;  
 KM autoimmune pulmonary inflammation; organ rejection; inflammation;  
 KM CLASP.  
 OS Homo sapiens.  
 XX  
 PN WO200231117-A2.  
 PD 18-APR-2002.  
 XX  
 PD 15-OCT-2001; 2001WO-US32202.  
 XX  
 PR 13-OCT-2000; 2000US-0687837.  
 XX  
 PA (ARBO-) ARBOR VITA CORP.  
 PA (GARM/) GARMAN J D.  
 PA (CAND/) CANDIA A F.  
 XX  
 PI Lu PS;  
 XX  
 DR WPI; 2002-416861/44.  
 DR N-PSDB; ABK84965.  
 XX  
 PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating  
 PT an immune response, and for treating multiple sclerosis, rheumatoid  
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,  
 PT and sepsis  
 PT  
 PS Disclosure; Figure 2; 245pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide (I) comprising an amino  
 CC acid sequence that has 90 % sequence identity to one of the human  
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2c, 2e)  
 CC sequences (PS). (I) is useful for identifying a compound or agent that  
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 CC inhibiting a immune response in a subject. A pharmaceutical composition  
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing  
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 CC the autoimmune disease is caused or exacerbated by increased activity  
 CC of Th1 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or  
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2  
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2  
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or  
 CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, Disgeorge syndrome, lymphopenia, thrombocytopenia, or  
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or  
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,  
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,

CC	endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
CC	inflammation. CLASP-2 can be used to treat anaphylaxis or
CC	hyperresistivity to an antigenic molecules, organ rejection or graft-
CC	versus-most disease (GVHD) and inflammation. ABG61670-ABG61708
CC	represent cadherin-like asymmetry protein (CLASP) sequences of the
CC	invention.
XX	
SQ	Sequence 1353 AA;
Query Match	62.0%; Score 6784.5; DB 23; Length 1353;
Best Local Similarity	92.7%; Pred. No. 0;
Matches 1353;	Conservative 1; Mismatches 3; Indels 101; Gaps 3;
QY	701 VLHHNHNPFEYDEIKELPTQLHEKHHLLTPFHVSCDSSKSGTKRDRVETQVYSWL 760
DB	1 VLHHNHNPFEYDEIKELPTQLHEKHHLLTPFHVSCDSSKSGTKRDRVETQVYSWL 60
QY	761 PLKDGKRVVTSSEHIVSANLPSGYLGYOELGKGRHYGPEIKWVDGKPLKISTHLVST 820
DB	61 PLKDGKRVVTSSEHIVSANLPSGYLGYOELGKGRHYGPEIKWVDGKPLKISTHLVST 120
QY	821 VYTDQHLNHFPOYCOKTSAGALGNELVKYLSLHMEGHVMAFLPTILNQLFRVLT 880
DB	121 VYTDQHLNHFPOYCOKTSAGALGNELVKYLSLHMEGHVMAFLPTILNQLFRVLT 180
QY	881 RATOEEVAVNVTXVLIHVVAQCHEEGLESLSYVYKAYKABPYVASEYKTVHEELTKSM 940
DB	181 RATOEEVAVNVTXVLIHVVAQCHEEGLESLSYVYKAYKABPYVASEYKTVHEELTKSM 240
QY	941 TTILKASADFLISNKKLTKTSWFFDVLKISMAOHLIENSKVKLLRNQRPAPSYHAAVEV 1000
DB	241 TTILKASADFLISNKKLTKTSWFFDVLKISMAOHLIENSKVKLLRNQRPAPSYHAAVEV 300
QY	1001 VNNLMHITQKRFNDPEASKNANSLAVFKRCFTMDRGFVFKOINNYISCAPDQPKT 1060
DB	301 VNNLMHITQKRFNDPEASKNANSLAVFKRCFTMDRGFVFKOINNYISCAPDQPKT 360
QY	1061 LFEYKEEFLRVVNCNHHYIPLNLPMPFGKRIQRYODLOLDVSLTDFCRNHFVGLLR 1120
DB	361 LFEYKEEFLRVVNCNHHYIPLNLPMPFGKRIQRYODLOLDVSLTDFCRNHFVGLLR 420
QY	1121 EVGTALQERREVRLLIISVKNLLIKGSPDDRYASGHQARIATVLPFLGLIENVORI 1180
DB	421 EVGTALQERREVRLLIISVKNLLIKGSPDDRYASGHQARIATVLPFLGLIENVORI 480
QY	1181 NVRDVSPFPVNAQMTVKDSLALPAINPLVTPQKSGTLDNSLHKDLGASIGASPYTTS 1240
DB	481 NVRDVSPFPVNAQMTVKDSLALPAINPLVTPQKSGTLDNSLHKDLGASIGASPYTTS 540
QY	1241 TPINISVRNADSRGSLISTDSGNSLPERNSEKSNLSIDKIQOSSLTGNVVRCDKLQDSEI 1300
DB	541 TPINISVRNADSRGSLISTDSGNSLPERNSEKSNLSIDKIQOSSLTGNVVRCDKLQDSEI 600
QY	1301 KSLIMCFYLTKMSMDALFTYNNKASTSELDFFITSVCLHOFQYMGKRYIARNQEGE 1354
DB	601 KSLIMCFYLTKMSMDALFTYNNKASTSELDFFITSVCLHOFQYMGKRYIARNQEGE 660
QY	1355 -----RTGMMHARLOQLGSLDNLSTPHNSYGHSAADVHLQSLLEANTA 1397
DB	661 GPTVHDRKQOTLPVSKNRRTGMHARLOQLGSLDNLSTPHNSYGHSAADVHLQSLLEANTA 720
QY	1398 TEVCLTALDPLSLFTLAFKQQLADHGHNPMLKGVVDVYLCPLOKQSEETALKNVETALR 1457
DB	721 TEVCLTALDPLSLFTLAFKQQLADHGHNPMLKGVVDVYLCPLOKQSEETALKNVETALR 780
QY	1458 SLIYKFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTYKSGF 1517
DB	781 SLIYKFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTYKSGF 840
QY	1518 VRTHLQVITISVQSLIDVYIGIGTRFOQSLIINNCCANDRLIKHTSFSSDVYDLTKRIR 1577
DB	841 VRTHLQVITISVQSLIDVYIGIGTRFOQSLIINNCCANDRLIKHTSFSSDVYDLTKRIR 900

QY	1578 TVLMATAQWKEHENDPEMLVDLOYSLAKSVASTPELRKTWLDNSVARIHVKNQDLSSEAMC 1637
DB	901 TVLMATAQWKEHENDPEMLVDLOYSLAKSVASTPELRKTWLDNSVARIHVKNQDLSSEAMC 960
QY	1638 YVHTVTAIAVEYITREKAVQWEPPLPHSHSACLRSGVFRQGTARTVRVITPNIDEAS 1697
DB	961 YVHTVTAIAVEYITREK-----GVFRQGTARTVRVITPNIDEAS 997
QY	1698 NMEDVGMQDVHFNENEVLMELBOCADGLMKAEYELIADIYKLIPIYEKRDPERLAML 1757
DB	998 NMEDVGMQDVHFNENEVLMELBOCADGLMKAEYELIADIYKLIPIYEKRD----- 1051
QY	1758 YDTHRAVSKYTVNHSGRRLGTYFRVAFPGQAQYQFTDSETVDEGFEDDEGKEYIY 1817
DB	1052 -----FEDEDEGKEYIY 1062
QY	1818 KEPKLTPLSEISQRLKLYSDKFGSENVKMIODSKVNPCKDLSKAYIQTHTVHPFDE 1877
DB	1063 KEPKLTPLSEISQRLKLYSDKFGSENVKMIODSKVNPCKDLSKAYIQTHTVHPFDE 1122
QY	1878 KELQERKTEFERSHNIRRFPMFPPTQTKQGGVEBOCKRTTILTAIHCFPYVKRIPV 1937
DB	1123 KELQERKTEFERSHNIRRFPMFPPTQTKQGGVEBOCKRTTILTAIHCFPYVKRIPV 1182
QY	1938 MYQHTDNLPIEVALIDENSKKVAELROCSSAEVDMIKLOLQGSVSQVNAAGLAYAR 1997
DB	1183 MYQHTDNLPIEVALIDENSKKVAELROCSSAEVDMIKLOLQGSVSQVNAAGLAYAR 1242
QY	1998 AFLDNTNTRKYPDNKVYLKEVFRQVFAACGQALAVNERLLKEDLEFOEEMKANYREMA 2057
DB	1243 AFLDNTNTRKYPDNKVYLKEVFRQVFAACGQALAVNERLLKEDLEFOEEMKANYREMA 1302
QY	2058 KELSEIMEOICPLEKTSVLPNSLHIFNAISGTPSTSMVGMCTSSSSSV 2107
DB	1303 KELSEIMEOICPLEKTSVLPNSLHIFNAISGTPSTSMVGMCTSSSSSV 1352
RESULT 12	
ABG61672	ID
ABG61672	standard; Protein; 1353 AA.
ABG61672;	
13-AUG-2002	(first entry)
XX	
DE	Cadherin-like asymmetry protein (CLASP) isoform #1.
XX	
KW	Human; autoimmune disease; haematopoietic disorder; Digesorge syndrome;
KW	blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;
KW	ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
KW	chromocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KW	haemolytic anemia; multiple sclerosis; rheumatoid arthritis; lupus;
KW	endometriosis; autoimmune thyroiditis; anaphylaxis; hyperresistivity;
KW	autoimmune pulmonary inflammation; organ rejection; inflammation;
KW	CLASP.
OS	Homo sapiens.
XX	
PN	WO200231117-A2.
XX	
PD	18-APR-2002.
XX	
PF	15-OCT-2001; 2001WO-US32202.
XX	
PR	13-OCT-2000; 2000US-0687837.
XX	
PA	(ARBO-) ARBOR VITA CORP.
PA	(GARM/) GARM J D.
PA	(CAND/) CANDIA A F.
XX	
PI	Lu PS;
XX	
DR	WPI; 2002-416861/44.

DR N-PSDB; ABK84966.

PT New human cathenin-1-like asymmetry proteins) (CLASP)-2 for modulating  
PT an immune response, and for treating multiple sclerosis, rheumatoid  
PT arthritis, endometriosis, lupus, autoimmune chryciditis, septic shock  
PT and sepsis -

PS Disclosure; Figure 3B; 245pp; English.

The invention relates to an isolated polypeptide (I) comprising an amino acid sequence that has 90 % sequence identity to one of the human cadherin-1-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E) sequences (BS). (I) is useful for identifying a compound or agent that binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for detecting a CLASP-2 polypeptide in a sample. (II) is useful for inhibiting a immune response in a subject. A pharmaceutical composition comprising a nucleic acid encoding (I), or (II) is useful for preventing or treating a CLASP-2 mediated disease e.g., an autoimmune disease, where the autoimmune disease is caused or exacerbated by increased activity of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for detection or inhibition of CLASP-2 expression (e.g., antisense or ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2 polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-antibodies or are used as therapeutic polypeptides. The CLASP-2 polynucleotide or fragments can be used in diagnostics (e.g., as probes for CLASP-2 expression), as a lymphocyte marker and for therapeutic purposes. CLASP-2 polynucleotides can construct transgenic and knockout animals, e.g., for screening of CLASP-2 agonists and antagonists. Polynucleotides can screen for CLASP-2 agonists and antagonists. CLASP-2 polypeptides or polynucleotides can treat deficiencies or disorders of the immune system, by activating or inhibiting the activation, differentiation of immune cells and can treat or detect deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides or polynucleotides can increase differentiation and proliferation of haematopoietic cells, including the pluripotent stem cells to treat those disorders associated with a decrease in certain (or many) types of haematopoietic cells e.g., immunologic deficiency syndromes including blood protein disorders (e.g., agammaglobulinemia, dysgammaglobulinemia, ataxia telangiectasia, common variable immunodeficiency, Disgeorge syndrome, lymphopenia, thrombocytopenia, or haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia, Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus, endometriosis, autoimmune thyroiditis, and autoimmune pulmonary inflammation. CLASP-2 can be used to treat anaplasia or hyper-sensitivity to an antigenic molecules, organ rejection or graft-versus-host disease (GVHD) and inflammation. AB561570-AB561708 represent cadherin-1-like asymmetry protein (CLASP) sequences of the invention.

**SQ**      **Sequence**      **1353 AA;**

Query Match	62.0%;	Score 6784.5;	DB 23;	Length 1353;
Best Local Similarity	92.7%;	Pred. No. 0;		
Matches 1345; Conservative	1;	Mismatches	3;	Indels 101; Gaps 3

Qy	701	VLHHQNEPFXDEIKELPTQHEKHHLLTFPHNSCDSSKSGSTKRPVVEQVGSYL	760
Db	1	VLHHQNEPFXDEIKELPTQHEKHHLLTFPHNSCNSKSGSTKRPVETQVGSYL	60
Qy	761	PLLDGRVVTSEQHI PVSANLPBGYLGYOELCMGRHYGPEIKMVDGKRLKISTHLYST	820
Db	61	PLLDGRVVTSEQHI PVSANLPBGYLGYOELCMGRHYGPEIKMVDGKRLKISTHLYST	120
Qy	821	VYTDQOHLPFQYQCKTBSGQALGNELVKYKLSLHMEGVMTLAFLEPTLNLQFLVLT	880
Db	121	VYTDQOHLPFQYQCKTBSGQALGNELVKYKLSLHMEGVMTLAFLEPTLNLQFLVLT	180
Qy	881	RATGEVAVNVTRVLIHHVAOCHEGLBESHLRSYKYVYKARPYVASEKTYHEELTSGM	940
Db	181	RATGEVAVNVTRVLIHHVAOCHEGLBESHLRSYKYVYKARPYVASEKTYHEELTSGM	240
Qy	941	TTILKPSADFLTSNLLKYSWFFDVLIKSMAQHLIENSKVCLLRNQRPPASVHAIVETV	1000

Db	241	TTIIKPSADPLTISKRLRYSMFFFDVLIKSMAQULIENSKYKULRNQRPSPSYHHAETV	300
Qy	1001	VNMIMPHITQKFRDNPEASKNVANSIAVFIKRCFTFMDRGCVFKOINNYISCFAPGDBKT	1066
Db	301	VNMIMPHITQKFGDNPEASKVANSIAVFIKRCFTFMDRGCVFKOINNYISCFAPGDBKT	360
Qy	1061	LFEXKPEFLRVVCHENHYIPLNLMFPGKRIQRYOOLQDYSLTDEFCRHHPLVGLLR	1120
Db	361	LFEXKPEFLRVVCHENHYIPLNLMFPGKRIQRYOOLQDYSLTDEFCRHHPLVGLLR	420
Qy	1121	EVGALAOEFREVRILIASVLKMLIKISFDDRVSRSRQARIATLYLPLFGLIENYORI	1180
Db	421	EVGALAOEFREVRILIASVLKMLIKISFDDRVSRSRQARIATLYLPLFGLIENYORI	480
Qy	1181	NVRDVSFPVNAGMTVDESLALPAVNPVLTPOKGSITLDSNLHDLGALSIGIASPYTTS	1240
Db	481	NVRDVSFPVNAGMTVDESLALPAVNPVLTPOKGSITLDSNLHDLGALSIGIASPYTTS	540
Qy	1241	TPNINSVRNADRSGLISTDSGNSLPERNSKSNLSDKQOOSTLGNSVVRCDKLDOSBI	1300
Db	541	TPNINSVRNADRSGLISTDSGNSLPERNSKSNLSDKQOOSTLGNSVVRCDKLDOSBI	600
Qy	1301	KSILMCFYLITKMSDDELFTYNNKASTSELMDEFTISEVCLHOFQVYKRYIARNOGL	1355
Db	601	KSILMCFYLITKMSDDELFTYNNKASTSELMDEFTISEVCLHOFQVYKRYIARNOGL	660
Qy	1355	RTGMMHRLIOLQGLSDLSLTFNHSYGHSDADVHLQSLLENIA	1397
Db	661	GPIYHDKSQCLPYSRRRTGMMHRLIOLQGLSDLSLTFNHSYGHSDADVHLQSLLENIA	720
Qy	1398	TEVCLTALDLSLFTLAPKQIOLADHGNPLMKKVPDYVLCFLQKHOSETLKNVFTALR	1457
Db	721	TEVCLTALDLSLFTLAPKQIOLADHGNPLMKKVPDYVLCFLQKHOSETLKNVFTALR	780
Qy	1458	SLIYKFPSTPEEGADMCALCYEILICCNKSLSITBASQLLYELMRNPDVTKGSF	1517
Db	781	SLIYKFPSTPEEGADMCALCYEILICCNKSLSITBASQLLYELMRNPDVTKGSF	840
Qy	1518	VRTHLQVYISVSQIADVVGIGGRPOOSLSIINNCSNDRILIGHTSPSSDVOLTKRIR	1577
Db	841	VRTHLQVYISVSQIADVVGIGGRPOOSLSIINNCSNDRILIGHTSPSSDVOLTKRIR	900
Qy	1578	TVMATAQMKHEHDPMLVDLOYSLAKSYASTELRKTMWDSMARIHVKNGLISEAAC	1637
Db	901	TVMATAQMKHEHDPMLVDLOYSLAKSYASTELRKTMWDSMARIHVKNGLISEAAC	960
Qy	1638	YVHTATLVAELITRKEAVQWEPULLPISHSACLRSKGVRCQCTAPRVTTPNIDBAS	1697
Db	961	YVHTATLVAELITRKEAVQWEPULLPISHSACLRSKGVRCQCTAPRVTTPNIDBAS	997
Qy	1698	MMEBVGQODVFNPNVDVLMELLEOCADGIMKARELIADYIKLIIPIYEXKRPDERLAHL	1757
Db	998	MMEBVGQODVFNPNVDVLMELLEOCADGIMKARELIADYIKLIIPIYEXKRPDE	1051
Qy	1758	YDTLHRAVSKYTEVWHSGRLLCTYFRVAFFGQAAOYQFTDSETVDEGFFDEDEGKEYIY	1817
Db	1052	YDTLHRAVSKYTEVWHSGRLLCTYFRVAFFGQAAOYQFTDSETVDEGFFDEDEGKEYIY	1066
Qy	1818	KEPULPLSEISQRLKLYSDKFGSENVKMIQDSGVNPKOJDSKXAYIQTWHVIPFDE	1877
Db	1063	KEPULPLSEISQRLKLYSDKFGSENVKMIQDSGVNPKOJDSKXAYIQTWHVIPFDE	1122
Qy	1878	KELOERTTEPBRSHNIRPFEMFEMFTYTGKRGQGVBOCRRRTLLTJHCPYVKKRIPV	1937
Db	1123	KELOERTTEPBRSHNIRPFEMFEMFTYTGKRGQGVBOCRRRTLLTJHCPYVKKRIPV	1188
Qy	1938	MYOHHITLNPJEVVIDEMSKKVAELROLCSAEVDMIKLOJLQGSYSVQVNAEPYLAAR	1997
Db	1183	MYOHHITLNPJEVVIDEMSKKVAELROLCSAEVDMIKLOJLQGSYSVQVNAEPYLAAR	1244
Qy	1998	AFIDDTNTKRYPDNKVULKEVFRQPYVACGQALAVNERLLIKEDQLEVOEBEMKANREMA	2057

DB 1243 AFLDDTNTKRYYPDNKVKLLKEVERQVEACGQALAVNERLIKEDQLEVOEMKANYRENA 1302  
 QY 2058 KELSEIMHEQICPLEEKTSVLPNSLHIFPAISGTPSTWVGHTSSSVY 2107  
 DB 1303 KELSEIMHEQICPLEEKTSVLPNSLHIFPAISGTPSTWVGHTSSSVY 1352

RESULT 13  
 ABG61686  
 ABG61686 standard; Protein; 1353 AA.  
 AC ABG61686;  
 AC  
 DT 13-AUG-2002 (first entry)  
 DE Cadherin-like asymmetry protein (CLASP) isoform.  
 KW Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;  
 KW blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;  
 KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;  
 KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;  
 KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;  
 KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;  
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;  
 KW CLASP.  
 XX Homo sapiens.  
 OS  
 XX MO200231117-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001MO-US32202.  
 XX  
 PR 13-OCT-2000; 2000US-0687837.  
 XX  
 PA (ARBO-) ARBOR VITA CORP.  
 PA (GARW/) GARMAN J D.  
 PA (CAND/) CANDIA A F.  
 XX  
 PI Lu PS;  
 XX  
 DR WPI; 2002-416861/44.  
 DR N-PSDB; ABK84973.  
 PT  
 PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating  
 an immune response, and for treating multiple sclerosis, rheumatoid  
 arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,  
 and sepsis -  
 PT  
 PS Example 4; Figure 6A; 245p; English.  
 XX  
 CC The invention relates to an isolated polypeptide (I) comprising an amino  
 acid sequence that has 90 % sequence identity to one of the human  
 cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2B, 2C, 2E)  
 sequences (PS). (I) is useful for identifying a compound or agent that  
 binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 inhibiting a immune response in a subject. A pharmaceutical composition  
 comprising a nucleic acid encoding (I), or (II) is useful for preventing  
 or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 the autoimmune disease is caused or exacerbated by increased activity  
 of T11 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 detection or inhibition of CLASP-2 expression (e.g., antisense or  
 ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 antibodies or are used as therapeutic polypeptides. The CLASP-2  
 polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2  
 polynucleotides can screen for CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or

CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptide  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or  
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or  
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,  
 CC endometriosis, autoimmune thyroiditis, rheumatoid arthritis, lupus,  
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or  
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-  
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708  
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the  
 CC invention.  
 XX  
 SQ Sequence 1353 AA;  
 Query Match 62.0%; Score 6784.5; DB 23; Length 1353;  
 Best Local Similarity 92.7%; Pred. No. 0;  
 Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLAHHQNEPFYDEIKELPTOLHEKHLILTFPHVSCNNSKSGTKRKDVETQVGSGL 760  
 DB 1 VLAHHQNEPFDEIKELPTOLHEKHLILTFPHVSCNNSKSGTKRKDVETQVGSGL 60  
 QY 761 PLKKGRRVTSBOHIVSANTPSGYLGYOELGMRHGYBEIKWVDGKPLKISTHLVST 820  
 DB 61 PLKKGRRVTSBOHIVSANTPSGYLGYOELGMRHGYBEIKWVDGKPLKISTHLVST 120  
 QY 821 VYTOPQHLNPFQYQKTESGQALGNELVYKSLHAMEGHVMAFLPTLLNQFRVLT 880  
 DB 121 VYTOPQHLNPFQYQKTESGQALGNELVYKSLHAMEGHVMAFLPTLLNQFRVLT 180  
 QY 881 RATOEVAVNVRVYIIHVAQCHEGLSHLSRYKYVYKAPYVASEYKTVHEELTSM 940  
 DB 181 RATOEVAVNVRVYIIHVAQCHEGLSHLSRYKYVYKAPYVASEYKTVHEELTSM 240  
 QY 941 TTILKPSADFLTSNKLKYSWFFFDVLKSAQHLIENSXYKLLNQFPASVTHAVETV 1000  
 DB 241 TTILKPSADFLTSNKLKYSWFFFDVLKSAQHLIENSXYKLLNQFPASVTHAVETV 300  
 QY 1001 VNMMLPHITQKRRDPKPSKANHSLAVYTRCFPFMRGPFVKQINNYISGFAGDPRKT 1060  
 DB 301 VNMMLPHITQKRRDPKPSKANHSLAVYTRCFPFMRGPFVKQINNYISGFAGDPRKT 360  
 QY 1061 LPEYKFEFLRVVYCNHEHYIPLNLPMPFGKGRIGRQDQLDVSITDEFGRNHFVGLLR 1120  
 DB 361 LPEYKFEFLRVVYCNHEHYIPLNLPMPFGKGRIGRQDQLDVSITDEFGRNHFVGLLR 420  
 QY 1121 EYGTALQEFREYRLAIVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVORI 1180  
 DB 421 EYGTALQEFREYRLAIVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVORI 480  
 QY 1181 NVRDVSPPPVNAAGMVKVKBESLAPVAVNPLVPQKSTLDNSLHKDLGAIGSAPYPTS 1240  
 DB 481 NVRDVSPPPVNAAGMVKVKBESLAPVAVNPLVPQKSTLDNSLHKDLGAIGSAPYPTS 540  
 QY 1241 TPNIIVSVNABSRGSLITDSGNSLPEKNSEKNSLDHGOOSTLGNSVVRCDKLDQSEI 1300  
 DB 541 TPNIIVSVNABSRGSLITDSGNSLPEKNSEKNSLDHGOOSTLGNSVVRCDKLDQSEI 600  
 QY 1301 KSLMWCPLIYIKSMSDALFTYWNKASTSELMDFETISVCLHQFYNGKXYIA----- 1354  
 DB 601 KSLMWCPLIYIKSMSDALFTYWNKASTSELMDFETISVCLHQFYNGKXYIANQEG 660  
 QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDAVYHOSLLEANIA 1397  
 DB 661 GPVHDRKSQTLPVSRNRRTGMMHARLQOLGSLDNLSTFNHSGHSDAVYHOSLLEANIA 720

QY 1398 TEVCLTALDTLSLFTLAFKNOGLADHGNPLMKKVEDVYLCELOKHQSESTALKNVFTALR 1457  
 DB 721 TEVCLTALDTLSLFTLAFKNOGLADHGNPLMKKVEDVYLCELOKHQSESTALKNVFTALR 780  
 QY 1458 SLIYKPFSTFYGRADMCAALCYELIKCNSKLSSTRTASQULYFLMNNNDYGGKSF 1517  
 DB 781 SLIYKPFSTFYGRADMCAALCYELIKCNSKLSSTRTASQULYFLMNNNDYGGKSF 840  
 QY 1518 VATHLQVITSVSQLADVVGIGGTRFQOSLSIINNCANSERLIKHSPSPSDVKDLTKRIR 1577  
 DB 841 VATHLQVITSVSQLADVVGIGGTRFQOSLSIINNCANSERLIKHSPSPSDVKDLTKRIR 900  
 QY 1578 TVLMAVTAQWKEHNDPEMLVDLOYSLAKSYASTPELAKTLWDSMARIHVNGDLSAANC 1637  
 DB 901 TVLMAVTAQWKEHNDPEMLVDLOYSLAKSYASTPELAKTLWDSMARIHVNGDLSAANC 960  
 QY 1638 YVHTVALVAEYLTRKNAVOMEPPLPHSHSACLRSGVFPQCGTAFFVTTPNIDEAS 1697  
 DB 961 YVHTVALVAEYLTRKNAVOMEPPLPHSHSACLRSGVFPQCGTAFFVTTPNIDEAS 997  
 QY 1698 RMEDYGMDOVHFNEEDVLMELBECADGLMKAREVELIADYIKLIPIYKRRDPERLANTL 1757  
 DB 998 RMEDYGMDOVHFNEEDVLMELBECADGLMKAREVELIADYIKLIPIYKRRDPERLANTL 1051  
 QY 1758 YDTLHRAVSKVTEVMSGRLLGTYPFVAFGQAQYQFPTDESDVEGFEDEDEGKEYIY 1817  
 DB 1052 -----FEDDEGKEYIY 1062  
 QY 1818 KEPKLTPLSEISQRLIKLYSDYFGSENNVMIDSGKNPKDIDSKAYIQTHTVIFPDE 1877  
 DB 1063 KEPKLTPLSEISQRLIKLYSDYFGSENNVMIDSGKNPKDIDSKAYIQTHTVIFPDE 1122  
 QY 1878 KEOERKTEFBSHNRIRFMPFQNGKROGVEOCCRRITLAIHCFFYKRIY 1937  
 DB 1123 KEOERKTEFBSHNRIRFMPFQNGKROGVEOCCRRITLAIHCFFYKRIY 1182  
 QY 1938 MYQHTDINPIEVAIDEMSKVAELRQLCSSAEVDMIKQLKQGSVSQVNAAGPLAYAR 1997  
 DB 1183 MYQHTDINPIEVAIDEMSKVAELRQLCSSAEVDMIKQLKQGSVSQVNAAGPLAYAR 1242  
 QY 1998 AFLDITNTKRYPDNKVKLLKEYFRQVVEACGQALAVNEELIKEDOLEYEEKMANYREMA 2057  
 DB 1243 AFLDITNTKRYPDNKVKLLKEYFRQVVEACGQALAVNEELIKEDOLEYEEKMANYREMA 1302  
 QY 2058 KELSHIMEQICPLEKTSVLPNSLHIFNAISGTPSTNVHGMTSSSVY 2107  
 DB 1303 KELSHIMEQICPLEKTSVLPNSLHIFNAISGTPSTNVHGMTSSSVY 1352

RESULT 14  
 ABG61679  
 ID ABG61679 standard; Protein; 1353 AA.  
 AC ABG61679;  
 DT 13-AUG-2002 (first entry)  
 DE Cadherin-like asymmetry protein (CLASP) isoform #8.  
 XX Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;  
 XX blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;  
 XX ataxia telangiectasia; common variable immunodeficiency; lymphopenia;  
 XX thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;  
 XX haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;  
 XX endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;  
 XX autoimmune pulmonary inflammation; organ rejection; inflammation;  
 XX CLASP.  
 OS Homo sapiens.  
 XX  
 XX WO200231117-A2.  
 XX

PD 18-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001WO-US32202.  
 XX  
 PR 13-OCT-2000; 2000US-0687837.  
 XX  
 PA (ARBO-) ARBOR VITA CORP.  
 XX (GARM/) GARMAN J D.  
 PA (CAND/) CANDIA A F.  
 XX  
 PI Lu PS;  
 XX  
 DR WPI; 2002-416861/44.  
 XX  
 PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating  
 PT an immune response, and for treating multiple sclerosis, rheumatoid  
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,  
 PT and sepsis  
 XX  
 PS Disclosure; Figure 5A; 245pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide (I) comprising an amino  
 CC acid sequence that has 90 % sequence identity to one of the human  
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)  
 CC sequences (PS). (I) is useful for identifying a compound or agent that  
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for  
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for  
 CC inhibiting an immune response in a subject. A pharmaceutical composition  
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing  
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where  
 CC the autoimmune disease is caused or exacerbated by increased activity  
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for  
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or  
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2  
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-  
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2  
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes  
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic  
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout  
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2  
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.  
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or  
 CC disorders of the immune system, by activating or inhibiting the  
 CC activation, differentiation of immune cells and can treat or detect  
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides  
 CC or polynucleotides can increase differentiation and proliferation of  
 CC haematopoietic cells, including the pluripotent stem cells to treat those  
 CC disorders associated with a decrease in certain (or many) types of  
 CC haematopoietic cells e.g., immunologic deficiency syndromes including  
 CC blood protein disorders (e.g., agammaglobulinemia,  
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable  
 CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or  
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or  
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,  
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,  
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary  
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or  
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-  
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708  
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the  
 CC invention.  
 XX  
 SQ Sequence 1353 AA;  
 Query Match 62.0%; Score 6782.5; DB 23; Length 1353;  
 Best Local Similarity 92.6%; Pred. No. 0;  
 Matches 1324; Conservative 2; Mismatches 3; Indels 101; Gaps 3;  
 QY 701 VLHHNPFYDEKIELEPTOLHEKHLLTFPHYSCNSKSGTKRQDVETQVGYMWL 760  
 DB 1 VLHHNPFYDEKIELEPTOLHEKHLLTFPHYSCNSKSGTKRQDVETQVGYMWL 60  
 QY 761 PLKXGRVVTSEOHIPVANSPLSGYGLQELGMGRHYGPRIKWDGKRLIKISTHLVST 820



Dd	61	PLKXGRVVTSGHIFVSAHNPBSGLGQELGMGHVPELIKMDGGRPLAKISHLVST	120
Oy	821	VYTDODHLNFPQYCKTESGAQALGNEIVKYLKSLHAMEGHVIAFLPTLLNDLFRVLT	880
Dd	121	VYTOQOHLNFPQYCKTESGAQALGNEIVKYLKSLHAMEGHVIAFLPTLLNDLFRVLT	180
Oy	881	RATOEVAWVWTRVIVHVAOCHEGEGESHLSRYKYVYKAEPPYASKEYVHEELTMSM	940
Dd	181	RATOEVAWVWTRVIVHVAOCHEGEGESHLSRYKYVYKAEPPYASKEYVHEELTMSM	240
Oy	941	TTILKPSADFLTSNKLKXSMFFDVLIKSAOHLIENSKYKLYLNORPPASYHHAETV	1000
Oy	241	TTILKPSADFLTSNKLKXSMFFDVLIKSAOHLIENSKYKLYLNORPPASYHHAETV	300
Dd	1001	VNMLMPTHOKFRODPEASKNANSHLAFIKRCFFPMRGVFNQINNYISCFAPGDKT	1060
Dd	301	VNMLMPTHOKFRODPEASKNANSHLAFIKRCFFPMRGVFNQINNYISCFAPGDKT	360
Oy	1061	LFEYKFEELRVVCNNEHYIPLULPMPFKGRIQORQDQLDLYSLTDEFCRNHFLVGLLR	1120
Dd	361	LFEYKFEELRVVCNNEHYIPLULPMPFKGRIQORQDQLDLYSLTDEFCRNHFLVGLLR	420
Oy	1121	EVGTLAQEPRRYLAIIVLKNLTIKXSPDDRYSRSHQARIATLYLFGILLIENVORI	1180
Dd	421	EVGTLAQEPRRYLAIIVLKNLTIKXSPDDRYSRSHQARIATLYLFGILLIENVORI	480
Oy	1181	NVRDVSPPVNAGMTVKBESLAPVNPVLVPQKSTLDSNLKDLGALISGASPYTS	1240
Dd	481	NVRDVSPPVNAGMTVKBESLAPVNPVLVPQKSTLDSNLKDLGALISGASPYTS	540
Oy	1241	TPNINSVANADSRGSLISTDGSNLPERNSEKSNSLDRHQOSTLGNSVVRCDKLDQSEI	1300
Dd	541	TPNINSVANADSRGSLISTDGSNLPERNSEKSNSLDRHQOSTLGNSVVRCDKLDQSEI	600
Oy	1301	KSILMCFYLIIXSMNSDDALFTYWNKASUSSELMDFITSEVCLHOCQYNGKRYIABNOEGL	1354
Dd	601	KSILMCFYLIIXSMNSDDALFTYWNKASUSSELMDFITSEVCLHOCQYNGKRYIABNOEGL	660
Oy	1355	-----RTGMMHARLQOLGSLDLSNLTPNHSYGSHDADVLHQSLLEANIA	1397
Dd	661	GPIVHDRKSQTLPVGRNRTGMMHARLQOLGSLDLSNLTPNHSYGSHDADVLHQSLLEANIA	720
Oy	1398	TEVCITADLTSLFTLAFKNOILLADHGNNPLMKKYFVUYLCFLQKHSETLKXVFTLKR	1457
Dd	721	TEVCITADLTSLFTLAFKNOILLADHGNNPLMKKYFVUYLCFLQKHSETLKXVFTLKR	780
Oy	1458	SLIYKPESTFEGRADMCAALCYEILKCNCSLISIRFEASOLYFLKRRNPFDTYGKXSF	1517
Dd	781	SLIYKPESTFEGRADMCAALCYEILKCNCSLISIRFEASOLYFLKRRNPFDTYGKXSF	840
Oy	1518	VRTHLOVIVISQALIVAVGIGTRFQOQSLIINNCAUSDRLIKHTSFSSDVOLTKRIR	1577
Dd	841	VRTHLOVIVISQALIVAVGIGTRFQOQSLIINNCAUSDRLIKHTSFSSDVOLTKRIR	900
Oy	1578	TVLMTATAMKEHENDPEMLVLDQVSLASVYSTPELRKTMVLDMSARHVNKGDISEAMC	1637
Dd	901	TVLMTATAMKEHENDPEMLVLDQVSLASVYSTPELRKTMVLDMSARHVNKGDISEAMC	960
Oy	1638	VYHVTALVAEYLITREAVOWMEPPLPHSHSACLRSRGVFRQGTARFVITPINDEBAS	1697
Dd	961	VYHVTALVAEYLITR-----GVFRQGTARFVITPINDEBAS	997
Oy	1698	MMEDVGMDDVHFNENEVIMELLBOCADGIMKAKERYLINDIYKLIPIYEKRDPERLAHL	1757
Dd	998	MMEDVGMDDVHFNENEVIMELLBOCADGIMKAKERYLINDIYKLIPIYEKRD-----	1051
Oy	1758	YDTLHRAVSKYTEVHNSGRLLIGTYFVAFQQAQOYFTDSETVDEGFPEDECKEYIY	1817
Dd	1052	-----FDEDECKEYIY	1062
Oy	1818	KEPKUTPLSEISQRLLLKYSDFGSENVKMIODSGKVPKOLDSKRAYIOVTHVIPFDE	1877

Db	1063	KEPLTLPJLSEISQGLKLTLYSPDKFGESENVKMIQDSGKVNPKDLDGKVAIVIQWTHVIPFDE	11222
Qy	1878	KELOERKTEEFRRSHNIRFPFEMFPFOTQGRQGSVEEQCRRTLLTIAHCPYPYKKRI PV	19377
Db	1123	KELOERKTEEFRRSHNIRFPFEMFPFOTQGRQGSVEEQCRRTLLTIAHCPYPYKKRI PV	11822
Qy	1938	MYQHHTLNPFEVIAIDEMSKVAELRQLCSSAEVDMLKLOLKGVSQVQNAAPLVAAR	19977
Db	1183	MYQHHTLNPFEVIAIDEMSKVAELRQLCSSAEVDMLKLOLKGVSQVQNAAPLVAAR	12424
Qy	1998	AFLDITNTKRYPDNKVKYLKEVFPFQYVACQALAVNERLIKEDQLEYOEEMKANRYEMA	20577
Db	1243	AFLDITNTKRYPDNKVKYLKEVFPFQYVACQALAVNERLIKEDQLEYOEEMKANRYEMA	13020
Qy	2058	KELSBINEHOICPLSEKTSVLPNSLHIFNAISGPTSTWYHGMSSSSV 2107	
Db	1303	KELSBINEHOICPLSEKTSVLPNSLHIFNAISGPTSTWYHGMSSSSV 1352	
RESULT 15			
ABG61687			
ID	ABG61687	standard; Protein; 1353 AA.	
XX	ABG61687;		
AC			
XX	13-AUG-2002	(first entry)	
XX			
DE	Cadherin-like asymmetry protein (CLASP) isoform #15.		
XX			
XX	Human; autoimmune disease; haematopoietic disorder; Digestive syndrome;		
KW	blood protein disorder; agammaglobulinaemia; dysgammaglobulinaemia;		
KW	ataxia telangiectasia; common variable immunodeficiency; lymphopenia;		
KW	thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;		
KW	haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;		
KW	autoimmunity; autoimmune thyroiditis; anaphylaxis; hypersensitivity;		
KW	autoimmune pulmonary inflammation; organ rejection; inflammation;		
KW	CLASP.		
OS	Homo sapiens.		
XX			
PN	WO200231117-A2.		
XX			
PD	18-APR-2002.		
XX			
PF	15-OCT-2001; 2001WO-US32202.		
XX			
PR	13-OCT-2000; 2000US-0687837.		
XX			
PA	(ARBO-) ARBOR VITA CORP.		
PA	(GARM/) GARMAN J D.		
PA	(CAND/) CANDIA A F.		
XX			
PI	Lu PS;		
XX			
DR	WPI; 2002-416861/44.		
XX			
PT	New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating		
PT	an immune response, and for treating multiple sclerosis, rheumatoid		
PT	arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,		
PT	and sepsis		
XX			
PS	Example 4; Figure 6B; 245bp; English.		
XX			
CC	The invention relates to an isolated polypeptide (I) comprising an amino		
CC	acid sequence that has 90 % sequence identity to one of the human		
CC	cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)		
CC	sequences (PS). (I) is useful for identifying a compound or agent that		
CC	binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for		
CC	detecting a CLASP-2 polypeptide in a sample. (II) is useful for		
CC	inhibiting a immune response in a subject. A pharmaceutical composition		
CC	comprising a nucleic acid encoding (I), or (II) is useful for preventing		
CC	or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where		
CC	the autoimmune disease is caused or exacerbated by increased activity		



of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for detection or inhibition of CLASP-2 expression (e.g., antisense or ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2 polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-2 antibodies or are used as therapeutic polypeptides. The CLASP-2 polynucleotide or fragments can be used in diagnostics (e.g., as probes for CLASP-2 expression), as a lymphocyte marker and for therapeutic purposes. CLASP-2 polynucleotides can construct transgenic and knockout animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2 polynucleotides can screen for CLASP-2 agonists and antagonists. CLASP-2 polypeptides or polynucleotides can treat deficiencies or disorders of the immune system, by activating or inhibiting the activation, differentiation of immune cells and can treat or detect deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides or polynucleotides can increase differentiation and proliferation of haematopoietic cells, including the pluripotent stem cells to treat those disorders associated with a decrease in certain (or many) types of haematopoietic cells e.g., immunologic deficiency syndromes including blood protein disorders (e.g., agammaglobulinaemia, dysgammaglobulinaemia, ataxia telangiectasia, common variable immunodeficiency, DiGeorge syndrome, lymphopenia, thrombocytopenia, or haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia, Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus, endometriosis, autoimmune thyroiditis, and autoimmune pulmonary inflammation. CLASP-2 can be used to treat anaphylaxis or hypersensitivity to an antigenic molecule, organ rejection or graft-versus-host disease (GVHD) and inflammation. ABG61670-ABG61708 represent cadherin-like asymmetry protein (CLASP) sequences of the invention.

Sequence 1353 AA;

Query Match 62.0%; Score 6782.5; DB 23; Length 1353;  
Beet Local Similarity 92.6%; Pred. No. 0;  
Matches 1344; Conservative 2; Mismatches 3; Indels 101; Gaps 3;

701 VLNHHNPEFYDEIKIELPTQLHEKHNLLTFPHVSCDNSKSGSTKCRDVEYQVSWL 760  
1 VLNHHNPEFYDEIKIELPTQLHEKHNLLTFPHVSCDNSKSGSTKCRVETQVQYSWL 60  
761 PLLKGRVVTSSQHTIPVSNLPSGUYCYOELGMRHRYPEIKWDSGKPLKISTHLVST 820  
61 PLLKGRVVTSSQHTIPVSNLPSGUYCYOELGMRHRYPEIKWDSGKPLKISTHLVST 120  
821 VTTDOHNLNPFYOXCKTSSGAOLGNELVKYLKSHAMESHVMIAFLPTIINQLFRVLT 880  
121 VTTDOHNLNPFYOXCKTSSGAOLGNELVKYLKSHAMESHVMIAFLPTIINQLFRVLT 180  
881 RATOEBVANVTRVIIHVVAQCHEBGLSHLSYVYKAYKAPYVASEKYTHEELTKSM 940  
181 RATOEBVANVTRVIIHVVAQCHEBGLSHLSYVYKAYKAPYVASEKYTHEELTKSM 240  
941 TTIKPSADFTLSNKLKTSWPPFDVLIKSMOHLIENSK/KLIRNQRPAHYHAAETV 1000  
241 TTIKPSADFTLSNKLKTSWPPFDVLIKSMOHLIENSK/KLIRNQRPAHYHAAETV 300  
1001 VMLMPLHTOKFRDNPDEASKNANSLAVPIKRCFTFMDGCVPKOINNTISCPAREDPKT 1060  
301 VMLMPLHTOKFRDNPDEASKNANSLAVPIKRCFTFMDGCVPKOINNTISCPAREDPKT 360  
1061 LFEYKFEFLRVVNCNEHYIPLNLPMPFGKRIQRYODLODLSLDEFCRNHFLVGLLR 1120  
361 LFEYKFEFLRVVNCNEHYIPLNLPMPFGKRIQRYODLODLSLDEFCRNHFLVGLLR 420  
1121 EYGTALQERREYRLAISTVLKULLIGSDRYASHSHQARIATVLPFGLLIENVQRI 1180  
421 EYGTALQERREYRLAISTVLKULLIGSDRYASHSHQARIATVLPFGLLIENVQRI 480  
1181 NVRDVSPPFVNAGMTWKDESLLPANNPLVTPQKSTLDNSLHKDLGAIISGIASPYTTS 1240  
481 NVRDVSPPFVNAGMTWKDESLLPANNPLVTPQKSTLDNSLHKDLGAIISGIASPYTTS 540

1241 TPINISVYNADSRGSLISTDSGNSLPERNSEKNSLDNRHQOSTLGSVVRCDKLDQSEI 1300  
541 TPINISVYNADSRGSLISTDSGNSLPERNSEKNSLDNRHQOSTLGSVVRCDKLDQSEI 600  
1301 KSLMCFLYILKSNDDALFTYWNKASTSELMDFITSEVCLHQOYNGKRYIA----- 1354  
601 KSLMCFLYILKSNDDALFTYWNKASTSELMDFITSEVCLHQOYNGKRYIA----- 1354  
1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDAVLHOSLLEANA 1397  
661 GPYHDRKSQTLPVSRNRTGMMHARLQOLGSLDNLSTFNHSGHSDAVLHOSLLEANA 720  
1398 TEVCITLADTLSTFLAFAKQQLADHGNPLMKKVFYDVLCTFQHQSEALAKVFTALR 1457  
721 TEVCITLADTLSTFLAFAKQQLADHGNPLMKKVFYDVLCTFQHQSEALAKVFTALR 780  
1458 SLIYFPGSTFEGRADMCALCYEILKCNKSLSSIRFASOLVFLMRNFDYTGKSP 1517  
781 SLIYFPGSTFEGRADMCALCYEILKCNKSLSSIRFASOLVFLMRNFDYTGKSP 840  
1518 VRTHLQVLIISVQLADVVGIGTRFQOGLSTINNCANSRLIKHTSPSSDYKDLTKRIR 1577  
841 VRTHLQVLIISVQLADVVGIGTRFQOGLSTINNCANSRLIKHTSPSSDYKDLTKRIR 900  
1578 TVLMATTAQMKHEHNDPEMLVDLYQSLAKSVASTPELRTKTLDSMARIHVKGNDLSEAMC 1637  
901 TVLMATTAQMKHEHNDPEMLVDLYQSLAKSVASTPELRTKTLDSMARIHVKGNDLSEAMC 960  
1638 YVHTALVAEYITKREAVQMBPPLLPHSHSACLRSRGCVFRQGTARVITPNIDEAS 1697  
961 YVHTALVAEYITKREAVQMBPPLLPHSHSACLRSRGCVFRQGTARVITPNIDEAS 997  
1698 NMEDVGMQDVHNEBVNLELBOCADGMLKARVELADIYKLIIPTEKRDPERLML 1757  
998 NMEDVGMQDVHNEBVNLELBOCADGMLKARVELADIYKLIIPTEKRDPERLML 1051  
1758 YDTLHRAVSKTEVWMSGRRLGTIFYRVAFGQAQOYFTDSETVESFPEDEDEKEYIY 1817  
1052 -----FDEDEDEKEYIY 1062  
1818 KEPKLTPLSEISORLLKLYSDKFGSENVKMIODSGKVPKOLDSKRAYIIOVTHVIPFDE 1877  
1063 KEPKLTPLSEISORLLKLYSDKFGSENVKMIODSGKVPKOLDSKRAYIIOVTHVIPFDE 1122  
1878 KELORKEFEPSHNIIRPFEMPTOTGKXQOGVEBOCKRRTILTALHCFPVYKRIPV 1937  
1123 KELORKEFEPSHNIIRPFEMPTOTGKXQOGVEBOCKRRTILTALHCFPVYKRIPV 1182  
1938 MYOHTDNLNPIEVAIDEMSKVAELRQCSSAEVDMIKLOLKLQGSVSVQVNAAGPLAAR 1997  
1183 MYOHTDNLNPIEVAIDEMSKVAELRQCSSAEVDMIKLOLKLQGSVSVQVNAAGPLAAR 1242  
1998 AFLDNTKRYPDNKKVLLKEVFRQFVACQOALAVNERLKEQDLRYOEEMKANYREMA 2057  
1243 AFLDNTKRYPDNKKVLLKEVFRQFVACQOALAVNERLKEQDLRYOEEMKANYREMA 1302  
2058 KELSIIMEQICPLEKTSVLPNSLHIFNAISGPTSTNMHGMSSSSV 2107  
1303 KELSIIMEQICPLEKTSVLPNSLHIFNAISGPTSTNMHGMSSSSV 1352

Search completed: July 14, 2003, 18:15:13  
Job time : 84.051 secs

THIS PAGE BEGINS BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 23.6546 Seconds  
(without alignments)  
2620.804 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936  
Sequence: 1 MSQPLLPASAEFRKRRAL.....ISGTPSTWVHGHTSSSVV 2107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359.5	3.3	1865	1 US-08-588-985-2	Sequence 2, Appl
2	359.5	3.3	1865	1 US-08-971-988-2	Sequence 2, Appl
3	199	1.8	3248	1 US-08-353-700-1	Sequence 1, Appl
4	199	1.8	3248	5 PCT-US95-16216-1	Sequence 1, Appl
5	177	1.6	3056	1 US-08-508-836A-8	Sequence 8, Appl
6	177	1.6	3056	2 US-08-629-001A-3	Sequence 3, Appl
7	177	1.6	3056	4 US-08-642-274D-3	Sequence 3, Appl
8	177	1.6	3056	4 US-08-952-127-3	Sequence 3, Appl
9	177	1.6	3057	4 US-08-952-014C-3	Sequence 3, Appl
10	173.5	1.6	3057	4 US-09-360-416-3	Sequence 3, Appl
11	173	1.6	3056	4 US-09-360-416-2	Sequence 2, Appl
12	172	1.6	3056	2 US-08-874-266-2	Sequence 2, Appl
13	167	1.5	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
14	163	1.5	2285	4 US-09-308-375-2	Sequence 2, Appl
15	162	1.5	2482	1 US-08-328-254-6	Sequence 6, Appl
16	159	1.5	4536	4 US-09-180-422B-27	Sequence 27, Appl
17	158.5	1.4	1388	4 US-09-572-191-2	Sequence 2, Appl
18	158.5	1.4	1388	4 US-09-723-262-2	Sequence 2, Appl
19	158.5	1.4	1388	4 US-09-723-219-2	Sequence 2, Appl
20	157.5	1.4	2987	2 US-08-970-269A-29	Sequence 29, Appl
21	157.5	1.4	2987	4 US-09-407-562-29	Sequence 29, Appl
22	157.5	1.4	3959	2 US-08-970-269A-30	Sequence 30, Appl
23	157.5	1.4	3959	4 US-09-407-562-30	Sequence 30, Appl
24	156	1.4	2329	3 US-08-755-587-16	Sequence 16, Appl
25	155.5	1.4	1786	4 US-08-973-462-8	Sequence 8, Appl
26	153.5	1.4	3418	2 US-08-639-501-2	Sequence 2, Appl
27	153.5	1.4	3418	3 US-09-044-946-2	Sequence 2, Appl

28	153.5	1.4	3418	3 US-09-044-908-2	Sequence 2, Appl
29	152.5	1.4	1038	4 US-09-541-782-4	Sequence 4, Appl
30	152.5	1.4	1038	4 US-09-723-820-4	Sequence 4, Appl
31	152.5	1.4	3418	3 US-08-755-587-44	Sequence 44, Appl
32	151.5	1.4	3418	2 US-08-603-753D-4	Sequence 4, Appl
33	151.5	1.4	3418	4 US-09-099-753-4	Sequence 4, Appl
34	151.5	1.4	3418	4 US-08-986-106-4	Sequence 4, Appl
35	148.5	1.4	1211	4 US-09-134-001C-4820	Sequence 4820, Ap
36	147	1.3	3144	2 US-08-457-273B-42	Sequence 42, Appl
37	147	1.3	3144	3 US-08-556-419-21	Sequence 21, Appl
38	147	1.3	3144	4 US-09-041-886-15	Sequence 15, Appl
39	146.5	1.3	3079	5 PCT-US94-00198-4	Sequence 4, Appl
40	146	1.3	2818	4 US-09-542-331-2	Sequence 2, Appl
41	146	1.3	2818	4 US-09-510-791-2	Sequence 2, Appl
42	145.5	1.3	1447	4 US-09-376-330-17	Sequence 17, Appl
43	144	1.3	3144	1 US-08-246-982A-6	Sequence 6, Appl
44	144	1.3	3144	1 US-08-453-265-6	Sequence 6, Appl
45	143.5	1.3	988	3 US-08-851-843A-69	Sequence 69, Appl

## ALIGNMENTS

RESULT 1  
US-08-588-985-2  
; Sequence 2, Application US/08588985  
; Patent No. 5777094  
; GENERAL INFORMATION:  
; APPLICANT: MICHIOYUKI MATSUDA et al.  
; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/588,985  
; FILING DATE: January 19, 1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1865 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL:  
; ANTI-SENSE:  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE: spleen cell of homo sapiens  
; US-08-588-985-2  
Query Match 3.3%; Score 359.5; DB 1; Length 1865;  
Best Local Similarity 18.7%; Pred. No. 5,4e-21;

	Matches	393;	Conservative	345;	Mismatches	764;	Indels	605;	Gaps	99
Qy	136	ROLPNKVVYLDL	PNVHYVEEDVEVDKDEDAASGOKGI	-TKHGMLYKGNNSAISVTM	194					
Db	3	RVAVPTK	-REEKYGAFAVYDAR	-GADELSTQIGDTHVILHETEGM	-YTL	51				
Qy	195	RS-----	PKRFFHL	-----IQGDGSKYFEFLKDIQKEP	-----	KSGIFLGFLYG	235			
Db	52	RKSKKGI	FPASVYHLKEALVEGKQKHETVRPGDL	-----PLIOEVTTTLREWSTIRQOLYV	108					
Qy	236	VSFRN	-NKVRSFAFELKMODSSYL	-----LAADSEVEMEWITL	-----	NKLTQNFEA	286			
Db	109	QDNREMFSSV	RMITDL	-IEMRSQILSGTLPPDELEKELKKYTAKIDYGNRLDLDLYVR	167					
Qy	287	MOEKNGDSHEDDEOS	KLGSGSGGLDYLPELAK	-----SAREAEIKLKSSEKRV	336					
Db	168	-----	DE-----	DGNILD	-----	PELTSTISLFRAHMISKQVEERLOEKRSOK	206			
Qy	337	LFYLDPDQKDFS	-----SABEYKSFBEKFKGRI	LYKNDLSFNLQCCVAENEGETT	391					
Db	207	-----	QNDINRQAKFAATPSLALF	-----VNLK	-----	NVCKIGEDAR	241			
Qy	392	NVEPFTVLTSPD	IKYNRKISADHFVDLHNFVSROMIAITTSPPALMNGSGPEQSLRGL	451						
Db	242	-----	VMSLTDVPYSEKFI	SENYLVWSSSLPDI	-----	DLHLNLRVAF	282			
Qy	452	HEAAMQYKOGI	FSVTCSPHDFILARIEKVLQSGITHCABEYMKSSDSKQAKYKMA	511						
Db	283	TDLGSKDKREKIS	FPVC	-QIVRGGR	-----	LADNTRKLTGSL	321			
Qy	512	KQACORLQGYR	MPMAARTLEFKASGNLDKARSAIYRQ	-----DSNKLSDMDMLKLIAD	568					
Db	332	-----	RRPFGAAMDVTDI	INGKVDDEDKHFIFPQFVAGENPFLQTVINKVIAA	371					
Qy	569	FRKPEKMAKL	PVILGNLDTITDNVSSDPFRVYNNSSYIPTKQETCSKPI	FEVEBEFPYC	628					
Db	372	KEVNHKGGGL	WVTLKLLPEDIHQIKKEPRLVLD	-----RTTAVARKTG	414					
Qy	629	IPKHTQPYTIYNNH	LYVVPKYLYKUSQKSAFARNIAICIEFKQDESDQPLKCIYRP	688						
Db	415	FPEIIMPDDV	-RNDLYVTLVGQDP	-KSKSTAKKVEYTVASVYD	468					
Qy	669	GGPVFTSAPAAV	LHNH	-ONPEYDEIKIELPTQHEKNHLLTF	FHVSCDSSKSGSTYK	747				
Db	469	GAGDEALISEYKSV	IYYQVQRPWFETVKA	PIEDVNSHILFTRHSSQDS	521					
Qy	748	RDVUTQYGYSL	RLYLK	-DGRVYVSEOR	-----	IPSAULPSCYLQYOLG	792			
Db	522	KDKSEKIFAL	ARVVKLMRYDGTTLRGEHDL	LYYKAEAKLEBDATYLLSPSTKALEBKQ	581					
Qy	793	MGRHNGPEIKWVD	---GKPLKISTHNVSTVYTODCHLHNFQYCSQKTESQAQALGHEL	849						
Db	582	HSA	-TGKMSQISGCTISKDSFOJISTVLCSTGLQDNDLLGLKKRSNT	-----	SLL	632				
Qy	850	VKYLKSLHMBG	HVMAIRLPTILNQLFRVLTATQGE	---VAWVTVIILHVAQCSHEE	905					
Db	633	QONLEQLKKV	DGSEVVKFLODLYALFNIMENSSSEFTDLVDFALVFIIGLADRKFQ	692						
Qy	906	GLESILRSYV	KAYAKA	-----EPYV	ASQYKVHBEILTSMTTILKPSADPLTS	953				
Db	693	HFNPLVLEYIK	HGFATLATAUKTLTKVLKNYVDGAKRPGVNBOLUYAMKA	-LESIFKFTVR	751					
Qy	954	NKLTKYSWF	PVDLTKSMAOHLIENSKYKLLRNQFRASYNHAEVTVNNMLPHNTOKER	1013						
Db	752	SRIL	-----FRQLYENKGEADFVSLSLQFRS	-----	INDMSSMSDC	789				
Qy	1014	DNPEASKNANSH	LAVFIKRCPTFMDRGFVKQ	---INNYISCFAPGDPKTLFEKXFEFL	1068					
Db	790	-----	TVRKGAALKYLP	TIIVNDYKLVF	---DPKELSKMFTFEFI	825				
Qy	1070	RVVCHNEHYI	PLANLPMRFGKGIQR	-YODLOLDYS	-LTDFECNNHPLVGLLREVGTA	1126				
Db	826	-----	LVNM	-GLLTIOQLYCLIEIVHSDLTFTONCSR	-EILLRPMITDOLKYNL	871				

QY	1127	QEFBEVRLIAISVLKNNLLKISPDORASRSHOKARIIVLYPLFGLLEIENQIRINRVDV	1186
Db	872	ERQEBLEACC-----QLLSH-----	897
QY	1187	PEPVNAGTIVDESEIALPAVNPVLTPQKGSTLNDLHKDLGALSIGIASPYTSTPNINS	1246
Db	898	P-----TQHNVOJIM-----	915
QY	1247	VRNADSRGSLISTDSGNSLPERNSEKNSLDKHQOOSTLGNISVVRCDKLDOSEIKSLMC	1306
Db	916	-RTVISMG-----	934
QY	1307	FLVYLKSGSDALFTYNNKAST-----SELMDFTEISEVCLHQQVWGKVIATGMNHA	1361
Db	935	MTALLRQEBD---HYAHALLITFGKMRDVAVD---LMETFLMKRLLGKNVYPPFDWVI--	987
QY	1362	RLQOLGSLDNLSTFN--HSYGHSDADVYLHOSIL--EANNIATEVCLTALDLSLFLAKGNQ	1418
Db	968	-----MNVQNVPLRAINOY-----ADMLNKKFLDQANFELD-----	1020
QY	1419	LLADHGNPLMKVFDVYLCLQKHQSEYALKNVFLARSLI-----YKPEST	1466
Db	1021	-----LMNMYFLHVAVFLT--QESLOLENSSAKRAKILNKYGDMEKQIGFEIRDM	1069
QY	1467	FY---EGRADCALCTEYLKCSKXSLSRTESQOLYFLMRNPNPTGSKSVFRHLO	1523
Db	1070	WYNLQGHKIKTPEWGPBILMTLIPETELKATITPFDMMQCEHST--RSQOMENE	1127
QY	1524	VIVISQOLIDVVGIGGTRFOOSL--SIINNCANSRLIK-HTSFSDVYDLTKRI---	1576
Db	1128	II---TKLDHVEVGGRBOYKVLFDKILHCHKKHXYLAKTGSTFVKLVLRMLERLDY	1184
QY	1577	RTVLMATPQKHEHDEBMLVDLOYSLAKSY--ASTELKTYLDSMARLHVKNKGDISEA	1634
Db	1185	RTIM-----HDENKENRMSCTVYNLNFKEJEREBEMYIRLYUKLCDLHKECDNYTEA	1236
QY	1635	AMCVNHTALVAEYLTREKAVOMEPRLLPHSHASCCLR--SRGVSFPGGCTAPRVTIPNI	1692
Db	1237	AY-----TLHLAKLIK-----SEDCVNAHLTORDB-----YQATT---	1268
QY	1693	DEBASMEVDVGMQDVH--FNEDVLMELLEQCADGLMKARERYELADIYKLIIPYEKRRDF	1751
Db	1269	--QOGLKQOLVQOEIHHYFDKQKOME--EALALGKELAEQYE-----NEMFDY	1311
QY	1752	ERLHLVDTLHRAVSKYTEVMSGRLLGTFRVAFGQAAQIQFTDSEIDVSGFEDED	1811
Db	1312	EOLSELKKQAOAFENIVKYI---RKPDPYFAGVYGO-----GFPFLR	1353
QY	1812	GKEYIYKRPKTLPSIISOIRLLKLYSDKFSENVKJMDQSGKVNPKDL--DSKAYIIOVTH	1870
Db	1354	GKVITTYGKEYEREDDEARLLTOPPY--AEKKITTSPPG---DICKISPGOYICST	1406
QY	1871	VIPFED-----EKELQERKTEPERSHNIIRRFMEEMPTQOTGR--QSGVEBOCKKRIILTA	1924
Db	1407	VKPLRDLPRKXHRVVSQIVSFYVNVNVQRPBVSRRPRKGBKPDNDFAMMIERTIYTT	1466
QY	1925	IHCPRPYKKKIPVNYQHNHTDNLPRLEVAIDEM---SKVAELKOLGSAEVDIMIKLOLK	1980
Db	1467	AYKUPGLIRMEFVKSVMVEISPLENAIETMOLTNDKINSMVOOHLDPSPJPIPLSMELT	1526
QY	1981	QGSUVQVONAGPLVAYARFLDPTNTKAYPD--NKVYKLKEVFPQFVACQOALAVNRLI	2038
Db	1527	NGIYDPAVWGSFANYECAFTRDRLQHNPRANEKIEKYLKOLIMQOIFPLAEGIRIHDDKV	1586
QY	2039	KEDOLEYOEBEKANVREMAKELSEIMHBOICPLEKTSVLPNSLIHFNALSGTPTSTMVH	2098
Db	1587	TEALRPRHERNEACFOKQKEVEX-----EYGRVIMPSSU---DDBRGSRPRSMVR	1634
QY	2099	GMTSSSS 2105	
Db	1635	SFTMPSS 1641	

RESULT 2  
US-08-971-988-2  
Sequence 2, Application US/08971988  
Patent No. 5786461  
GENERAL INFORMATION:  
APPLICANT: MICHIOYUKI MATSUDA et al.  
TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/971,988  
FILING DATE: 17-NOV-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/588,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1865 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
HYPOTHEICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: spleen cell of homo sapiens  
US-08-971-988-2  
Query Match 3.3%; Score 359.5; DB 1; Length 1865;  
Best Local Similarity 18.7%; Pred. No. 5.4e-21;  
Matches 393; Conservative 345; Mismatches 764; Indels 605; Gaps 99;  
Db 136 ROLPNKRVKLDKLVHVEVEEDVDKEDDASISGQKGI-TKQGLYKGNMNSAISVTM 194  
3 RWVPTK-REEKGVAFYVVDAR-GADELSLQIGDTVHILETYEGV-YRG-----YTL 51  
195 RS-----FKRRPFH----IQGDGSYKFEFLDKLQKPE-----KGSIFLGLYNG 235  
Db 52 RKKSXKGIIPASTIHLEKAIIVEGKGHEVTPGDL--PLIOEVTTLREWSITWQLYV 108  
236 VSEFRN--NKVRRFAPFLKMDKSSYL--LAADSEVEMEWITIL-----NKILQLNFEAA 286  
Db 109 QGNREFRSVRHMIYDL-IEWRSQILSGTLPODELKELKKVYAKIDYGNRIIDLDLYVR 167  
287 MGEKNGSDHEDDEQSLGSGSGGLDSYBELAK-----SAREAEIKIKSSRYK 336  
Db 168 -----DE-----DGNILD--PELTSTISLFRABRIAKQVEERLQEEKSQK 206  
337 LFYLDPAQKLDPS-----SABPEVKSFEKPKGKRLVVCNDSFYLQCCVANESEGPTT 391  
Db 207 -----QNIDINROAKFPAATPSIALF-----VNLK-----NVVCKIGDAE--- 241

Qy 392 NVEPFVTLSLEFDIKYNRKISADPHVDLNRHSVRQMIATTSBALMGSGPETQSALRGIL 451  
Db 242 -----VLMSLYDVPESKFISENLYVRWSSGGLPKDI-----DELHNLRAVF 282  
Qy 452 HEAMQYRKQGISFTCPHPDIFLVARIKVLQSGITHCABPYMKSSSSKVAQKVLKRA 511  
Db 283 TDLGSKDKIKREKISFVC---QIVRVGRME-----LNDNTRKLTSGI----- 321  
Qy 512 KQACORLQGYRMPFMAARTLEKQASGNLDKVARSAIYRQ---DSNLSNDMDMLKLD 568  
Db 322 -----RRPFVAVMDVTDIINGKYDEDEKHFIPQPVAGBNPFLQVINKVIAA 371  
Qy 569 FKPEPKAKLPVILGNLDITIDNVSSDPENYVNSSYIPTKQFETSKPTIPFEVEFPVC 628  
Db 372 KEVNRKGGOLWTLTKLPDGIHQIKPEFLVD-----RTAVARKTG----- 414  
Qy 629 IPKHTQPTTYTNHLVYVPKYLKVDSQSPAKARIAICIEPKDSDEDSQPLKITYGR 688  
Db 415 FPEIIMPQDV-RNDIYVTLVQGDPD-KGSKTKAKRVEVTSVYD---BDGKRLHVI-FP 468  
Qy 689 GGPVFTRSAPAAVLHHH--ONPEFYDEIKLEPTOLHEKHLLLPFHVHSCNSKSGSYTK 747  
Db 469 GAGDEAISEKYSVITYQKQPMFETVVAALPDEDVNRSHLRFTRHSSQDS----- 521  
Qy 748 RDVETQVGYSWLPULK-DGRVVTSEOR-----IPVSANLPSCYLGYOBIG 792  
Db 522 KQKSEKIFALAPVVKLMRYDGTTLRGEHDLVYKAKAKKLEDAATYLSLPTKALBEKG 581  
Qy 793 MGRHYGPETKAVDG---GKPLKISTHLVSTVYTDQHLNFPQYCKTGESAOALGNEI 849  
Db 582 HSA-TGKSMOSIGSCITISKDSFQISTLVCTLTQNVLLGLKRSNT-----SLI 632  
Qy 850 VVYLSLHMBGHVWIAFLPTINQFRVLPRATOE---VAVNTVVIHVVAQCHEE 905  
Db 633 QGNLQKLVKVDGGEVVKFLDQTLDLALFNIMMENSESEFTDLVFPALVPIIGLIDRKFQ 692  
Qy 906 GLESHRSYVVKYAKA-----BPVY-ASEYKTVHEELTSMNTILKPSADFLTS 953  
Db 693 HNPVLEITYIKHFSATLAATYTLTKLVKNYDGAKEPQVNBQLYAKMA-LESIFKPIVR 751  
Qy 954 NKLKYSWFFPDVLIKSPAQHLIENSKYKLLRNQRFPAHYHNAVETVNMIMPHITOKFR 1013  
Db 752 SRIL-----FNQLYENKGEADFVESLQLPFRS-----INDMSSMSDQ-- 789  
Qy 1014 DNEPASKNANSLAFIRSCFTFMDRGFVPKQ---INNYISCFAPSGPKTLFEKPEEL 1069  
Db 790 -----TVRVGALKATYLPITVNDVXLPF---DPKELSKNFTPEI 825  
Qy 1070 RVVCHNHHYIPLNLPMPFGKGRIGR-YODLDLDS--LTDFECRNHPVGLLREVGTAI 1126  
Db 826 -----LNVFM--GLITIQKLYCLIEIYHSDLFQNDCK-BILLPMPTDQDKYHL 871  
Qy 1127 GEFREVRLLAISVLKNLLIKISFDDRVSASHQARIATVYLPFGLILENQRINRVDS 1186  
Db 872 ERQEDLEACC---QLLSH-----ILEVLYR---KQVG 897  
Qy 1187 PEPVNAQMTVDESLALPAAVNPVTPQKSTLDSNLHMDLGAISGIASPYTTSPPNINS 1246  
Db 898 P-----TORHVQIIM-----EKLRTVA----- 915  
Qy 1247 VNAADRSGLISTDSGNLPERNSEKSNLKHQOQSFTLGNSSVVRCDKLQSEIKSLMLC 1306  
Db 916 -RTVISMG-----RSESLIGNFVA-----C 934  
Qy 1307 FLYIYKSSDDELFTYNNKAST-----SELADFITISVCLHOFQYMGRIYIARGMMMA 1361  
Db 935 MTAIRQMED--YHAAHLITFGKMRDVAVDL-LMETPIMFNKILIGNVYPRPDWVI-- 987  
Qy 1362 RLQGLSLDNLGTFN--HSYGSADVYHQSL--BANATVEYCLALDTLSIFTLAPKQ 1418  
Db 988 -----NNMQNKKVFLRAIQY---ADMINKKFLDQANFELQ----- 1020  
Qy 1419 LLAHDGHNPLMKKVFDVYVLCFLQKHQSEETALKVNFALRSIL-----YKFPST 1466

Db 1021 -----LMNNYFHLAVAFLT--QESLQLENFSSAKRAKILNKYGDMRQIGFEIRD 1069  
 Qy 1467 FY---EGRADMCALCEIILKCNCSKSSIRTEASOLLYLMMNNPPTKSGFVATHQ 1523  
 Db 1070 WYMLQGHKIFPIBMPGPILEMTLIPETELRKATITPFFPMOCCEPST--RSFQWENE 1127  
 Qy 1524 VIISVQLIADVVGIGSTRFQOSY--SIINNCANSRLIK-HTSFSSVDYKDTTKRI-- 1576  
 Db 1128 IT---TLDHVEVSGRDEQYKLPFDKILLEHCKHKLKLTGETFYKLVVRLMRLD 1184  
 Qy 1577 RTVLMATAQMKHEHNDPEMLVDLOYSLAKSY--ASTPELKTWLDMSARIHVXGDLSEA 1634  
 Db 1185 RTIM-----HDENKEMNSCTVYANLYKFEIEREMRYLYKLCIDHKECDMYTEA 1236  
 Qy 1635 AMCVHTALVAEYLTRKEAVOMEPRLLPHSHSACLR--SRGVPFGQCTARVITPNT 1692  
 Db 1237 AY-----TLHLAKLLKM-----SEDCVAHLTORQJ-----YQATT-- 1268  
 Qy 1693 DEBASMEVDVGMODVH-FNEVDLMELLEOCADGLMKAREYELIADIYKLIIPYERKRD 1751  
 Db 1269 --QGOLKEQLYQELIHFDKGMWE--EALALGKELAEQYE-----NEMFDY 1311  
 Qy 1752 ERLAHLVDTLHRAYSKVTYVMSGRRLGTTFYVAFFGQAQYQFTDSETVGEPFEDD 1811  
 Db 1312 EQLSELKKQAQEFENIVKVI-----RPKPDYFAGVYQ-----GPPFLR 1353  
 Qy 1812 GKRYIKERKLTLSLSEISQRLKLYSDKPSSENVKMIQDSGXNPKDL-DSKAYIQVTH 1870  
 Db 1354 GKPFITRGKYEYEPDEEARLLTOPFN--AEKKKTTSPG--DDIKNSPGQYIQCT 1406  
 Qy 1871 VDPFD-----EKELOERKTEPERSHNIRREMEPFTQGR--QGVBEQCKRRTILYA 1924  
 Db 1407 VKRKLDPKFNHPVSEQIVSFYRVNEVQGFESRPIRKGKXKPNDBFAMWIERITTYT 1466  
 Qy 1925 IHCFPVYKRIPLYMTOHTDNLNPLEVAIDEM---SKCAELQOLCSSAEVDMIKIQLK 1980  
 Db 1467 AYKLPGLRFEVYSVYVMEISPLENAIETMQLTNDKINSMVOQHLDPSLPINPLSM 1526  
 Qy 1981 QGSVSQVYVNAAPLAVARFLDENTRKYPD--NKVKLKEVPFQVAFBAGQALAVNERLI 2038  
 Db 1527 NGVDPVAVMGFANYEKAFFTDYLOEHPAHEKIEGLKOLIMQIPFLAEGIRIHGXV 1586  
 Qy 2039 KEDOLEYOEEMKANYEMAKELSEIMHEQICPLEEKTSVLPNSIHFIENASTPTSTMY 2098  
 Db 1587 TBLRPRHERMEACFQKKEKVEK-----EYGVRLMPSL--DDRGRSRRPSMVR 1634  
 Qy 2099 GMTSSSS 2105  
 Db 1635 SFTMPSS 1641

RESULT 3  
 US-08-353-700-1  
 Sequence 1, Application US/08353700  
 Patent No. 5599919

GENERAL INFORMATION:  
 APPLICANT: YEN, TIMOTHY J.  
 APPLICANT: RATTNER, JEROME B.  
 TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
 TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN.  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
 STREET: 1601 MARKET STREET, SUITE 720  
 CITY: PHILADELPHIA  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-2307  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL TYPE: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN

Query Match 1.8%; Score 189; DB 1; Length 3248;  
 Best Local Similarity 18.5%; Pred. No. 1e-06;  
 Matches 444; Conservative 342; Mismatches 828; Indels 700; Gaps 107;

Qy 94 CSTVPAKAEENAS-----LFTVECTITNSDWHLVNXYKYEDYSGEFRQLEN 140  
 Db 121 CSELEERSQQAQASADVSLNPCNTPQKIFTPPLTPSQ-----YSGSKYEDLEKRY-----N 172  
 Qy 141 KYVXLDKLPVHYVEUDEVDKEDDAISLGSQKG-----ITKH-----GWL 181  
 Db 173 K-----EVEERRGLEAEVKALQAKASQTLPOATNHRDIDARHOASSVFSQ 220  
 Qy 182 YGNMNSAISVTMRSEFKRFFHLIQDGSYFBEFLKDLQKPK-----GSTFL 230  
 Db 221 QKTSLSHSSNQRPPIRDF-----SASY-----FSGLEAVTPSSSTIQIGRDNASSGF 271  
 Qy 231 G-----FLYVSFNNKVRPAFELKMODSSYLLAADSEVEWETITLTKI--LQLN 282  
 Db 272 GNSSPHLLDQKAQNOELRNKINLELR-----LQGEHEKWKQ-----VKKFQELDQ 321  
 Qy 283 PEAAMQEKRGSHDDQSKLESGSGLDSTLPBLASAR---EALIKLS--ESRYKLF 338  
 Db 322 LEKAVEL-----LEKEV-----LTKCRDELVRTAQQYDQASTKYTALQKXKL 367  
 Qy 339 YLDPDAQKLDFASSA-----EPVYKSPERKGRILVKNCDLSPNLQCVVAENEGPTTVE 394  
 Db 368 TEDLSQGNOMESACSLQKIKKEKEKFOEEL-----SRQ 404  
 Qy 395 PFFVTLSLFDIKYNNKISADPHVDINHPSVQ---MIATTSPALNGSGPETGALRGI 450  
 Db 405 RSFQTLDEQCMKARLQELQDAKNMNVLAQELDKLTSVQQLENN----- 452  
 Qy 451 LHEAMQYFKQGISFVTCRPHDILVARI--EYVLOGSTHHC--AEPYMSSDSKVAQKV-- 507  
 Db 453 -----LEBPQKL-----CRAEQAFQASQI KENELRSMEMKKNLLKSHSEQAFAREVC 503  
 Qy 508 -----LKNAKQA-----CQRLGYRMPFAMAARTLPDQASGNDLKNARFSAIYRQDSNLSN 559  
 Db 504 HLEALKHIKQCLNSQNFAPAEEMKAKTSQETMLRDLQEKIN-----QQNSLTL 553  
 Qy 560 DDMKLADLFRK-----PEKAKLPVILGNLDITIDNVSPFPYVNSSYTPKQFETCS 614  
 Db 554 EKLKLAVALDLEKQDCQDLKKRHHIEQLNDKLSKTEKESKALLSALBKKEKEBELK 613  
 Qy 615 KTPITF-----EVEFVPCIRKHTQPTTYITNHLVYRK-----YIKYSQ 655  
 Db 614 EKKTLFSCWSENEKTLQMESEKENLQSKINHETCLTKTQIKSHBYNVRVITLEMRE 673

QY 656 KSPAKANITACIEFDSDEBDSOPKCIYGRPGGVFTTRSAFAVAHLHHNNEFYD--- 712  
 DB 674 NLSVEIRHNHNLVDSK-SVEVEFOKL-----AYMELOQKAEFFSDQH 714  
 QY 713 --EIK-----IELPQLEKHLHLLTFPHVSCDSSKSGTKKRDVVEQVGSMLPIL- 763  
 DB 715 QKEIEMCKLCTSQLOQVEDELEKLOL-----SNEIMEKDRCYQDILAEIESLRDL 767  
 QY 764 --KQGVNTSEOHIPVSAWLPBGYLGYOELGMGRHNGPEIHWVGKRPILKISTHVVSTV 821  
 DB 768 KSKDMSJLVNREDB-----QKSLNLPDOOPAMHNSFANITIGQSGMSEBSECELEA-- 818  
 QY 822 YVQDQH-----LHNEFYQYQKTESGAQALGNELVYKLSLHMEGHVNIAPLPTLNOQ 875  
 DB 819 ---DQSPKSAIIONVDSLEFSELSQKQNSDLOQCEBELVQIKGEIENIMKA--EQM 873  
 QY 876 FVLVTRATQBEAVANVTYIHH--VVAOCHEBESLEHSLASYKAVAKAPYVASEKTYH 933  
 DB 874 HOSPAVATISQRIKQOEDTSAHQNVAA-----TISALENTEKELOLNDKXETEQAEL- 927  
 QY 934 BELTSMTTILKPSADFLTSNKLKYSWFFPVLKISMAQHILNISKVTLRNORPASY 993  
 DB 928 QELK-----SNHLEBDSLKELOLSTLSLEKKENSITSLKRE----- 968  
 QY 994 HHAVENTVNMMLPHITQKFRDNEBASKNANHSIAVEIKCCTFMDRGVFKOINNYISCF 1053  
 DB 969 -----IEELTOENGTLKEINASINQKGNLQKSESFA-----NYIDBR 1007  
 QY 1054 ASGDPTLEFYKPEPRVVCNENHYIPLNMPFGKGRIOF-----YODLODYSLTD 1106  
 DB 1008 EKSISSLSDOYKOEKIL-----LOKCEBTGNAVEDLSQKXKAAQ 1047  
 QY 1107 EFCRNHFLVGLLREVTALQEFREVRLLAISVLKYL--IKHSFDDRVASRSHOARIAT 1164  
 DB 1048 E---KNSKLECLNEC-TSLCENR-----KNELEOLKEAF-----AKHQEFLTK 1088  
 QY 1165 LVLPLEFG-----LLIENVOIRINVDVSPPEVN--AG-----MTVKDESIALPAV 1206  
 DB 1089 L--AFAEBRNQNLMELEFTVQOALRSWMTDNONNKSSEAGLQKQIMTLKEE----- 1138  
 QY 1207 NBLVTPQKSGTLDNSLHK--DLGALISGLASPYTT-----STPNINSVRNADSRGS 1255  
 DB 1139 -----ONKQKQKENVNDLLOENBOLMKVMKTKHCQNLSESPINNSVAKERESEBN 1186  
 QY 1256 L-----ISTDSGNS-----LPERNSSEKNSLDHQOOSTLGN-S 1288  
 DB 1187 QCNFRQMLVEYKEISLDSTYNAQULVLEMLRNKELKLOESBEKEKCLQHELOTRGDLE 1246  
 QY 1289 VVRCDKLQDSBKSLIMCPL--YI-----LKSMSDPAFTYWNKASTSELMDFPTI 1337  
 DB 1247 TSNLOQMOQOEISGLKDCEDIDAEKYSISGPHLSSTQNDNNAHLOCSLOTTNKNLNELEKI 1306  
 QY 1338 SEVCLHOFQYMGKRYIARTGMNAR-----LOQJGSLDNLSTLPHNSYGHSDADVL 1387  
 DB 1307 CBI-----LOAEKELVTELDNRSECTATRTKMAEVEGCLNEVKIL-----NDOSGLL 1356  
 QY 1388 HOSLLE-----ANATEVCLTALD-----TIS-----LFTLAF 1415  
 DB 1357 HGBELVDIDIGGERGEOPNEOHPSLAPLBESNSYEHLLTISJKEVOMHMFALOEKFLSLOS 1416  
 QY 1416 KQOLLDAGHNPMLKKVPVYLCTLOKHOSETALKNVF--TALRS--LIYKFPSTFYE 1469  
 DB 1417 EKHILHDO-HCQMSKMSSE-----LOTYVDSLKAENLVSTYMLRNROGLVEMOUGLE 1470  
 QY 1470 GRADMCALCYELKCNKSLSTIRTEASOLYFLMRNPFYTYGKKSFPVTHLOVITISVS 1529  
 DB 1471 GLVPSLSSSCVP--DSSSLSSLDGS-----FYRALLQOTGDMSL-SNIEGAVSAN 1519  
 QY 1530 QLIADVVGIGTRFOOSLIINNCAUSDRLIKHTSSSDVKDITKIRVLMATAQMKH 1589  
 DB 1520 QCSVDEV-----FCSSLQTYVDSLKAENLVSTJLRNFOJDLVK----- 1558  
 QY 1590 ENDEPMLVDLOXSLAKSVASTPELRKTWLDMA-----RIHVKNGLS----- 1632

DB 1559 -----EMOJLEBGLVPSSLSSCVPPDSSLSLSIGDSSFYRALLEQTDGDKSLSNLEGVSA 1614  
 QY 1633 --EAMCVYHVAIVAEYITTRKEAVQWEPPLPHSHSACLRBSRGVFPQCGAFVITPN 1691  
 DB 1615 NQCSVDEVPCCSLQEBENLTRKETPSAPAKVEEELSLC-----EYHQ-----S 1658  
 QY 1692 IDEBASMEDVQMOVHFNEBDMLELQCADGLMKARVEL--IADYKLIIPYEKRR 1749  
 DB 1659 LEKLEBKESQSIMK--NKEI--QELFO-----LSSRQELDCLARKY-----LSENG 1704  
 QY 1750 DFERLAH--YDILHRAYSKTEVMA-----SGRLLIGTYFRAVAFQQA 1791  
 DB 1705 WQOKLSTVLEWESFLAEKQTEQLSLEVARLOGLDLSRSLG----- 1753  
 QY 1792 AQOYFTSEDTVEGFEEDE--CKEYIYKEPULPTLSELSORULKYSKDPSE-----NVK 1846  
 DB 1754 ----IDTEDALQGNHSCDISKEHTSETTEPTPHVDHQ-----ICKDAQODLNLDE 1803  
 QY 1847 MIDSGKYNP-----KDLDSKYAVIQVHVIPEPDEKELOERKTEPERSHNRPMPE 1899  
 DB 1804 KITETGAVKPTGECGSEGSPTNYE-----PGEEDKTQSSSECISELSTSGNALVP 1855  
 QY 1900 MPTQTGRQGGVEBOCKRRITLTAIHCFPPYKRIIPVMTQHTDNLNLEVAIDEMSKV 1959  
 DB 1856 MDFL-----GNQED-----IH-----NLQLRVETSNEVILRLHV-IEDRDKV 1893  
 QY 1960 A---ELAQOLGSAVDMIKLOLKQOQSVQVANAFLAYARAFDIDNTKRYPMNKYL 2015  
 DB 1894 ESLNEMKELDSKHLQEOVLMTKTEACTIELEKTVIGELKENSJLSE-KLEVPSCDHOL 1952  
 QY 2016 LKEVPROFEACQALAVN-ERLIKED-----OLAEYQEMKANYREMAELSEIMEQICP 2070  
 DB 1953 LQRV--ETSEGLNSLMEHNAKSSREDIGDNVAKNDSEKRFIDVENHLSIR----- 2004  
 QY 2071 LEEKTSVLNLSLHT 2084  
 DB 2005 -SEKASIBHEALYL 2017

RESULT 4  
 PCT-US95-16216-1  
 ; Sequence 1, Application PC/TUS9516216  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ven, Timothy J.  
 ; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
 ; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street Suite 720  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/16216  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/353,700  
 ; FILING DATE: 09-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Reed, Janet E.  
 ; REGISTRATION NUMBER: 36,252  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 PCT-US95-16216-1

Query Match 1.8%; Score 199; DB 5; Length 3248;  
 Best Local Similarity 18.5%; Pred. No. 1e-06;  
 Matches 424; Conservative 342; Mismatches 828; Indels 700; Gaps 107;

QY 94 CSTVPAKAEBAOS-----LPVTECTIKYNSDMHLYNKYEDYSEFQOLN 140  
 DB 121 CKSELRSSQQAASADVSLNCPNTPKIFTTPLTPSQO---YSGSKYEDLKKY----N 172  
 QY 141 KYVKLDKLPHVHYVEEDVEDDAISIGSKG-----ITKH-----GWL 181  
 DB 173 K-----EVEERKLEAEVAKLQAKKASOTLPQATMNRDIAHQASSVFSNQ 220  
 QY 182 YKGNMNSAISVTNRSFKRRPFHLIQDGSYKFEFLKDIQKEPK-----GSIFL 230  
 DB 221 QEKTPSHLSNSQRTPIRRDF-----SASY---FSGELEVTPRSSTLQIGKRDANSFF 271  
 QY 231 G-----FLYGVFRNNKVRFAFELKMDKSSYLLAADSEVMEWITLTKI--LQIN 282  
 DB 272 GSSSSPHLLDQLAQOELNKNKINLELR---LOGHEKEMKGQ---VNFQELQ 321  
 QY 283 FEAMQEKNGDSDHEDEQSKLEGSGGLDSYLPETAKSAR--BAIKLKS--ESRVLK 338  
 DB 322 LEAKAKEL-----IEKEV-----LNKCPDELVTTAQYDQASTKTALEQLK 367  
 QY 339 YLDPDAQKLDPSA---BPEVKSFEKFGKILVKNCLDSFNLQCCVAENEGPTTNYE 394  
 DB 368 TEDLSQORQASARSCLBOKIKEKEFOEBL-----SRQ 404  
 QY 395 PRFVTLSDPIKKNKISADFVNDLNFVSRO---MIATSPALNNGSGPETQSLRGI 450  
 DB 405 RSTQTLDOECIQKALTOELQAKMHNVLQAEILDKLTSVKQOLENN----- 452  
 QY 451 LHEAAMQPKGIFSVTCRHPDIFLVARI-EKVLQGSITHC--AEPKSSDSKVAQY- 507  
 DB 453 -----LEBRQKL-----CRAEQAFQASQIKENELRSMEMKKNLLSHSQKARBYC 503  
 QY 508 -----LKNAKQA---CQRLGQYMRPAMARFLFKDASGNLDKNARFSAIYRQDSNKL 559  
 DB 504 HLEAEIKNIKQCLNQSQNFABEKAKANTSOETMLRDLQEKIN-----QOENSLTL 553  
 QY 560 DMVLKLLADPRK-----PEKMAKLPYLGLDITINSSDPFNYNSSYIPKQEPETS 614  
 DB 554 EKTKLVAADLEKQDCSDLLKREHHIQQLNDKLSKTEKESKALLSALEKKKEVEELK 613  
 QY 615 KTIPTT-----EVEEFVPCIPKHTOPTYITNLUYYPK-----YLKYDSQ 655  
 DB 614 EKKTLFSCSKSEKSKLLTQMESEKENLQSKINHLFELCTQOQKSHENYERVTLEMDRE 673  
 QY 656 KSPFAKANIATICIEFSDSEEDSQPKCIYGRPGGFVTRSAFAVLLHHQNEFYD--- 712  
 DB 674 NLSVEIRNLHNVLDSK-SVEVETQKL-----AYMELOQKAEFSDQGH 714  
 QY 713 --EIK-----IELPQLHEKHLHLLTFHVSCDNSKSGSTKKRDVVEVOVGSWMLPL- 763  
 DB 715 QKEIEMKCKTSQLTQVEDLEHKLQL-----SNEIMDKRCYQDHLAEBSLRDL 767  
 QY 764 --KDGAVTISEQHI PVSANLPSSGYLQYQELGMRHNGPEIKWVDSKPLKISTHVLSTV 821  
 DB 768 KSKDABLVNEDH-----QBSLAFDQOPAMHHSFANITIGEQSGMPERSCRLEA-- 818  
 QY 822 YTDQDH-----LHNEFYQCKTESGAOLGNELVYKLSLHAMEGHVMIAPLPIIQL 875

DB 819 ---DQSPKSAILOQRVDLSLEFSLESQKQMSDLOKQCEBELVQIGKEIENLMKA--EQM 873  
 QY 876 FRVLTRAQOEBVAVNVTVIIH--VVAOCHEGLESLSRYVYKAYKAEPPVASYKVVH 933  
 DB 874 HQSVAETSQRIKSQEDTSAHQNVAAE---TLSALENKEKELQILNDKVEFOAEI- 927  
 QY 934 EELTKSMWTILKPSADFLTSNKLKYSWFFDVLIKSAOHLIENSKYKLLRNQRFPA 993  
 DB 928 QELKK-----SNHLEDSLKEQLLSETLSLEKEMSSIISLNKRE----- 968  
 QY 994 HHAETVNMMLPHITQKFRDNPPEASKNANSLAVFIKCTFEMDRGVFOQINNYISCF 1053  
 DB 969 -----IEELTOENGLTKEINASINQEKMLIQKSSFA-----NYIDER 1007  
 QY 1054 ACPDKTLFEYKFEELRVVCHNEHITPLNMPFKGIOR-----YODLQDLSLTD 1106  
 DB 1008 EKSISELSDQYKQELIL-----LQCEETGNAVEDLSQKTKAAQ 1047  
 QY 1107 EFCRNHPLVGLLRREVGTALQEFREVRLAISVLKNLL--IKHSFDDRVASRQARLAT 1164  
 DB 1048 E---KNSKLECLINBC-TSLCENR-----KNELEQLKEAF---AKEHQEFILK 1088  
 QY 1165 LYLPLFG-----LIENYQRIWVDVSPPPV---AG-----MTVDESLALPAV 1206  
 DB 1089 L---AFABERNQMLLELETVOALRSEMTDQNNNSKEAGGLKQEIWTLKEE----- 1138  
 QY 1207 NPLVTPQKSGTLDNSLHK---DLGAGISGISPYT-----STPINSVRNDRSGS 1255  
 DB 1139 -----QNMQKEVNDLLOENBOLMKWTKTKHCEONLSEPRNSYKESERBN 1186  
 QY 1256 L-----ISTDSGNS-----LPRNRSKNSLSDHQOOSTLGN-S 1288  
 DB 1187 QCNFRQMDLVEKELSLDSYNAQLVQLEAMLRNKLKLOESKEKECQJHELOITRGDLE 1246  
 QY 1288 VVRCDKLQOSEIKSLMLCPL-----YI-----LKSNSDDALFTYWNKASISLMDFTTI 1337  
 DB 1247 TSNLDQMSOEISGKDEIDAEKYGISGPHSLSTQSDNHNALQCSLQITWNKLNLEKI 1306  
 QY 1338 SEVCLHQYQWNGKRYARTGMMHAR-----LOOLGSLNLSLTFNHSGSDAVL 1387  
 DB 1307 CBI-----LOAEKELVTELDNRSECTITTRKMAEVEGKLNEVKIL---NDOSGL 1356  
 QY 1388 HQSLLE-----ANATEVCITALD-----TUS-----LFTLAF 1415  
 DB 1357 HGEIVEDIHPGFEGRQPRQHPVSLAPLDEBSNYEHLTSDQEVQMHFALEQKELSQS 1416  
 QY 1416 KQIILLADHGNPLMKKVPDYILCFLQKQSETALKNVF--TALRS-----LYKFPSTYE 1469  
 DB 1417 EHKIILHDO-HCQMSKMBE-----LQTVVDSLKENLVLTSLRNFOGDLVKEMQGLDEE 1470  
 QY 1470 GRADMCALCYEILKCNCKLSIRTEASQULLFLMRNNFDYTKGSEFVPTHLOVITISVS 1529  
 DB 1471 GLVPSLSSCVF---DSSSLSSLGDS---FYRLLEDTGMSL--SNLECAVAVAN 1519  
 QY 1530 QLIADVIGIGTRFOOSLSIINNCANSRLIKHTSFSSDVXDLTKRITVLMATAQMEH 1589  
 DB 1520 QCSVDEV-----FSSSLQTVVDSLAKENLVLTSLRNFOGDLVK----- 1558  
 QY 1590 ENDPMLVDLOYSLAKSYASTPELRYTLDSMA-----RHHVKNGLS----- 1632  
 DB 1559 -----EMOGLBERGLVPSLSSSCVPSDSSLSLIGSSFYRALLEQTDGMSLSSNLEGVSA 1614  
 QY 1633 --EAMCYHVITALVAEYLITRKEAQQWMEPRLPHSHASCLARRSGRVFPGOGCAFRVITPN 1691  
 DB 1615 NQCSVDVFCSSLOEENLTRKETPSAPAKGYBELSLC-----EYVQO-----S 1658  
 QY 1692 IDEBASMEDEVQMDVHNEEDVLMELLECADGLMKARVEL--IADYKLIIPYEKRR 1749  
 DB 1659 LEKLEEKNESSQIMK--NKEI--QELFQ-----LSSRQELDCLRKQY-----LSEBQ 1704  
 QY 1750 DFERLAHI---YDTLHRAYSKYTEVMH-----SGRLLIGTYFRVAFPGQA 1791



Db 1705 WOQKLTSTLWESKLAERKOTBOLSLLEVARLOGLULSSRLG----- 1753  
Qy 1792 AQOYFDSFDEGEFEDED-SKEYLYKEPKLTPLSEISQRLKLKYSDKFSSE---NYK 1846  
Db 1754 -----IDTDAIAGNRESCDISKEHSETERTPRHDVHQ-----ICDDAQQDNLDE 1803  
Qy 1847 MIQDSGRNP-----KOLDSKAVIQTTHVIPPEDKELOERTEFRSHNIRPFME 1899  
Db 1804 KLTETGAVFTECEGSEQSPDTNVE-----PREDKTOGSSECSISLSEFSGNALVP 1855  
Qy 1900 MEPTQGRKQGVBEQCKRRTILTAHCFPVYKRIPIVNYJHTDNLPEVAIDEMSKV 1959  
Db 1856 MDPL-----GNOED-----IH-----NLQRYKETSINENRLTHV-IEDRDKV 1893  
Qy 1960 A-----ELRQCSSAEVDMIKLOKLQGSVSVQVNAGPLAYARAFDDTTRKYPDNKVYL 2015  
Db 1894 EBLNEMWELDKLHGEVQWTKTBACIELEKIYELKENSJLSE-KLEFVSCDHQL 1952  
Qy 2016 LKEVFRQVYACGQALAVN-ERLIKED-----OLEYQEEKNANYREMAKELSEIMEHQICP 2070  
Db 1953 LQRV--ETSEGLNSDLEMHADKSRREDIGDNVAKVNDSWYERFLDVENELSRIR----- 2004  
Qy 2071 LEKTSVLPNSLHI 2084  
Db 2005 -SEKASIEHEALYL 2017

## RESULT 5

US-08-508-836A-8  
Sequence 8, Application US/0850836A  
Patent No. 5777093  
GENERAL INFORMATION:  
APPLICANT: Shilo, Yosef  
APPLICANT: Taglio, Danilo A.  
APPLICANT: Collins, Francis S.  
TITLE OF INVENTION: Ataxia-Telangiectasia Gene  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Reising, Echington, Barnard & Perry  
STREET: P. O. Box 4390  
CITY: Troy  
STATE: Michigan  
COUNTRY: US  
ZIP: 48069  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508, 836A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: P-313 (TAU)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 689-3500  
TELEFAX: (810) 689-4071  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3056 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-508-836A-8

Query Match 1.6%; Score 177; DB 1; Length 3056;  
Best Local Similarity 17.3%; Pred. No. 7e-05;  
Matches 435; Conservative 348; Mismatches 83C; Indels 918; Gaps 118;

Qy 66 LINDCRLMLFPYDDQTALILRQGRYICSTVPAKAEERAGSLFVTECIKTIVNS-DWHLV 124

Db 271 LINDSLKEVIT--ELFQLOI-----YIHPRKGAATQCKG-----AYESTKRSI 311  
Qy 125 NYKED-----YSGFEROLPNVYVLDKLPVHYVEYDERVDXDEDAASGS 170  
Db 312 LYNLYDLVNEISHIGSRKYSGGRNIAVKNELIELMADICHQV-----FREDTRSLFI 366  
Qy 171 QKGGITKGMWLYKGNMNSAIVTMRSPRRRFFHLIQLDGGSYKEFFLD-LQKEPKGSIF 229  
Db 367 SQSYTTQ-----RESSDYSVPCRRK-----IELG-----WEVIDHLQKSGNDPL 409  
Qy 220 LGFL-----YGVFRNNKV-----RRPAFLK-----MDKS 256  
Db 410 VPMLOIATOLISKYPASIPNCELSEPLMILSOLFPQHRGERTPYVRLTEVALCDPR 469  
Qy 257 SYLLAADSEVMEETITLNLKILQNFPAQOEKNGS-----HEDDQSK 303  
Db 470 SNLESSQKSDLLKLM--NKCITTFRGISSEBOQAEFGLGAIQGSILVEYDEREWK 525  
Qy 304 L-EGSG-----SGLDYLPFLAKSAREABIKLKSESRVYK- 336  
Db 526 LFTSACRPSCPAVOCCLTALITTSIVPGAVMGIEQNNCEVNRS-----PSLK-ESIMKW 579  
Qy 337 -LFT-LDPPAQKLDSSAEPEV--KSFEKFGKRIIVKCNDSLSENLQCCA- 383  
Db 580 LLFYQLEGD--LENSTEVPPIHSHNPHLVLEKILV-----SLTMKCKKAMNPFQGV 631  
Qy 384 -----ENBGGPTNVPFV--TLSLFDI-----KNRKISADPFVLDNHSV 424  
Db 632 ECEHHKKEKESISFEVELEFLQTFDMDPLTVRECGIERKQSSIGPSVHONKESLD 691  
Qy 425 ROMIATTSPALMGSGPETQS-----ALRGILHR-----A 454  
Db 692 RCLGLSGLQNLNNTSSRTNSETLVRCGRLLVGVLCGCTWGVIAEBAAYKSELFQKXNS 751  
Qy 455 AMQYPKQGISFVTCPHDIPLVARIKYLQGSITHCAPYKSSDSFVA-----Q 505  
Db 752 LMQAGBSITLPIKNTNEFRIGSLRNMWQ-LCTFCLSNCTKKS-PNKIASGFLRLTTS 809  
Qy 506 KYLKNAKAQCORLQY-RMPF-AMAAKTLFDDAGNLDMKARFSAIYQDSNKLSD--- 560  
Db 810 KLMDIADI CKSLASFIRKPPDRGEVESMEDTNGNL-----MEVEDOSSMNLFNDDPD 863  
Qy 561 -----DMLKL-----ADPRKEK 574  
Db 864 SSVSDANPESGOSTIGAINPLAEYLSKODLPDMLKFLCLCTVTAQTNTVSRADI 923  
Qy 575 MAKLPVILGNLDITIDNVSDFPNVNNSYIPTQPFETCSKTPITFE-----VEEF 625  
Db 924 RKLLMLID-----SSTLEPTSLHMLTMLKLELPGBEYPLPMBDV 966  
Qy 626 VPCIPKHTQPTIY-----TNHLVYTPKYL--KYDSQKS----- 657  
Db 967 LELKPLSNVCSLYRRDQCKTILMHVAVKMLGQSNMSEENTRDQOGLTVIGAFW 1026  
Qy 658 -----PAKARNIACIE-FKSDSEDSQPLKCIYGR--PGGVPTFSAPFALVHH 705  
Db 1027 HLTERRKTIYSVRNALVCLTLLBADPYSKMALINWKGDFPVNSVFTQ--FLDNNHQ 1084  
Qy 706 -----ONPEFY 711  
Db 1085 VRMLAASINRLFQDTKDSRLKALPLKLOTAFAENAYLKAQGMREMSHANSNPFL 1144  
Qy 712 DEIKLELPTQLHEKHHLLTFPHV--SGDSSKSSSTKRDVVEYQVGSWMLPKDGRV 769  
Db 1145 DEI-----YNRKSVALTLIAVVLSCS-----PICEKQALFALCKSVENGIL- 1185  
Qy 770 TSEOHIPVSA-NLPSGYLYQOELG--MGRHYG-PEIKVVDGSKPLKISTH-LVSTVYTO 824  
Db 1186 --EHLVAVKULEKYSSETGYRLDEPFMAHSHLDIVLEKMLAQDTBNYLSSEPPILNVT- 1242  
Qy 825 DOHLNFPQYQCKT-----ESGAQALGNELVYKLSLHA-MEGHVNIAPFLPTILN 873

```

Db 1243 --NIEDFYRSCYKLLPHLVIRSHFDEVSIAHQIOEDWKSLLTDCFPKILVNLFPYAY 1300
Qy 874 QLFVRLTRATNOEVAN-----NVTNIIHVNAQCHEEGES-- 909
Db 1301 EGTROSGMAQOETATKYDMLKSENLOKQIDHFTSNLPELVLELTLHEPANSSAS 1360
Qy 910 -----HLRSYVKAAYKAEPPYVAEYKTHEELTKSMITLKESADP 950
Db 1361 QSTDLCDGSGDDPAPNPPHPSHVIAKATFA--YIENCHKT-----KLKSLILELSSP- 1413
Qy 951 LHSNKLKTSWFFPDVLKISMAOHLIENSVKLLRNQRPASVHAHAEVTVNMMLEHTIQ 1010
Db 1414 -----SYQILALICEQAAETNNV--YKGRILIKIYH-----LFVSLLEKDI-- 1453
Qy 1011 KFRDNPASKNANHSALVFIKRCFTFMDRGFVKOINNTYISCPAGDPKTLREYKFEPL 1070
Db 1454 -----KSGLGAMAFVLRDVITYT-----LIHYINQRPSCMDVSLNS--FSLCCDLIS 1499
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLOLDYSLTDEFCRNHPLVGLLREVGOTALQEBR 1130
Db 1500 QVC-----QTAVTYCKDALENHL-----HYIVGTLI-----PLVYEGV 1532
Qy 1131 EYRLAISVVKULLIHSFPDDRYSRSHQRIATLYPLFGLLIENVQIRINRDSPPV 1190
Db 1533 EYQKQVLDLKYLVINDKNE-----NLV-----ITIKLDPFP- 1566
Qy 1191 NAGMTYKDESLALPAPNPLVTPQKSTLNSLHKOLLGAIISGIASPYTTSTPNINSVRA 1250
Db 1567 -----DHVVEKDL-----RITQOKIK 1582
Qy 1251 DSRGSLISTDGSNLSLPERNSEKSNLSLKHQOSSTLGSVVRCDKLDQSEIKSL--LMCP- 1308
Db 1583 YSRGPF-----SLBEINHPLSVS-----YDALVTLTLEGKDLRQL 1621
Qy 1309 YLKSNSDALFTYMKKASTSELMDFTTSEVCLHOFQYMKRYIARTG-----MMAR 1362
Db 1622 ELHKQOMVIM-----RASQDNPDQDIMVLYV--NLQLSKAIIHHTGEKYLEAVGSC 1674
Qy 1363 LQOLGSLD--NSLTFNHSYGHSDADVHQSILLEANIA--TEVCLTAL--DTL-----SL 1410
Db 1675 LGEVGIIDSTIAIHSKQASATKAL--KLEFEDKELQMTFIMLYTNLTVLEDQVAVRA 1732
Qy 1411 FTLAFKQULADHGHN--PLMKVFPDVLCLFQKHQSETA-----LKNVFTALRS- 1459
Db 1733 AVTCLKNILATKTHGSFWEIYKATTDPMLAYIQPRTSKKPLEVPRPKENPFEBLDDI 1792
Qy 1460 -----IYFESTIEGRADMCALCYELK--CNSKSSITTEASQLIYFLM 1505
Db 1793 NMIPLSENHDIWIKLTCAFLDSGCKEIL--QLKEMCEVKTDFCQTVLPYLHIDL 1850
Qy 1506 RANFDYTKKSPFRTHLOVYIISQILADVIGIGTRFQOSLSIINNCAUSDRLIKHTSF 1565
Db 1851 LODTNESW--RNLISHTVQGF--SCLNHSQTSRSTTPPAL 1889
Qy 1566 SSDVKD-----LTKRIITVLMATAQKHEHNPENMLVDQYSLAKSYASTPELAKTMD- 1619
Db 1890 DSESEHFFRCCLDKSQRTMLAVDVMRQKRS-----SGTIFNDAMWLD 1936
Qy 1620 ---SMARIHVKGDDSEAMCYVHTALV--AEYLTRKEAVQWEPRLPHSHSACLRSSG 1675
Db 1937 NYLEVAVK-----AOSCAAHFALLYAEIYADKKMDQO-----EKKS 1974
Qy 1676 GVPRQCTAFRVITPNIDEASMBDVGMODYHFNEDVLMEL-----LEQCADGIM 1726
Db 1975 LAFEBGOSTTI--SSLSEKSKETGISIQ-----DLLEIYRSIGEPDSLYGCGG-- 2024
Qy 1727 KAERYELADIYKLLIPIYEKRRDPERLAHLUYTLHRAVSKYEVMSGRRL----- 1778
Db 2025 -----KMLQPI--TRLRTYEHA--MMKCALVUTDLEIAIISSTRQAGIIQALON 2070
Qy 1779 LGTYFVAEFGAQAQYFTDSETDV-----GFEDDEGKEYIYKEPK 1821
Db 2071 LGLCHILSYVLKGLDYENKDWCPLEBELHAYOAMRMQMCHTSVSKVEGTS--YHESL 2128

```

```

Qy 1822 LTPLEISORLLKYSDKFGSENVKMIODSGVNPKDLSKYAVI----- 1866
Db 2129 YNALOSLADREHSTYESLTKARKEVEEMCK--RSLESYISLYPTLSRQAIGELESI 2185
Qy 1867 -----QVTHVIPFEDEKELOERKTEPERSHNIR--FWEMPTQTKRGQVEEOCK 1917
Db 2186 GELFSRSVTH-----RQSEVVIKQKHSQQLKDSDFSFQEPIMAL----- 2226
Qy 1918 RTTITLHCPFYVKKRIPWYQHHTDNLPIEVALIDSKYVABEOLCASAENVMIK-- 1975
Db 2227 -RTVI-----LEILMEKEMDNSQRECKIDILTKLVELSLIARTFKNTOLPER 2273
Qy 1976 --LOKLQGSVSVQVNAEPLAVARAF-----LDDTNTKRYPDNKYK 2014
Db 2274 AIFQIKQVNSVSCGSENBQLEAQVFMAKESQSLASTIKOMIKKLDASCANNPS----- 2329
Qy 2015 LKEVFRQFVEACGQALA-----VNERLIKEDOLEYQOEMKANY--REMAKEL 2060
Db 2330 -LKLTYTECLRVCGNWLAEETCLENPAVINQTYLEKAVENAGVYDESSDEL 2379

```

```

RESULT 6
US-08-629-001A-3
; Sequence 3, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; TITLE OF INVENTION: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290,00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-629-001A-3
Query Match 1.6%; Score 177; DB 2; Length 3056;
Best Local Similarity 17.3%; Pred. No. 7e-05;
Matches 435; Conservative 328; Mismatches 830; Indels 918; Gaps 118;
Qy 66 LINDCIREMLLPYDFQALILRQGRYICSTVPAKAEBAQSLFVTECIKTYNS--DWHLV 124
Db 271 LINDSLKEVII--ELFQLOI-----YIHHPKAGATQCKG-----AYESTKMSI 311

```

QY 125 NKYED-----YSGEPQLENKVKLDKLFVHYEVEDEVEDKEDDASLGS 170  
DB 312 LYNLDVLYNEISHIGSRGKSGGFENIYAKENILJELMADI CHOV-----FNEDETRSLERI 366  
QY 171 OKSGITTKGMLYKGMNNSAIVTWSFKRPFHLLQIGDGSYKFEELKD-LCKEPRGSIF 229  
DB 367 SOSYTTTO-----RESSDYVPCRRKK-----IELG-----WEYIKDHLKQSONDFL 409  
QY 230 LGFL-----YGVFPNNKV-----FRFAFELK-----MODKS 256  
DB 410 VPMLOIAQOLISKYRPSLPCNLSPLMLLSQLLPQQRHGERTPYLRCLTEVALCODR 469  
QY 257 SYLLAADSEVEEMEWITTLINKILQLEFAMQEKNGDS-----HEDDEQSK 303  
DB 470 SNLESOSKSDLIKLW-----NKIWCITFRGISSEQKQENFCLGAILQIGSLVEVREFFK 525  
QY 304 L-EGSG-----SGLSYLPBELAKSAREAITKLSBSRYV- 336  
DB 526 LFTGSACRSPCAVCCLTALITTSIVPGAVKMGIEQNMCEVNRS-----FSLK-BSIMKM 579  
QY 337 -L-FY-LDPPAQKLDSSAEPV--KSFEKFGKRIYKCNELSFNLQCCVA----- 383  
DB 580 LFPYQLEBG--LENSTEVPILHSNPHLYLEKILV-----SLTMKQKAMNFPQSV 631  
QY 384 -----ENEBGPTTVBPPV--TSLFDI-----KYNRKISADPHVDLHPSV 424  
DB 632 BCEHHKHKEBELSFSEVEELPLOTTPDKMDPLTIYRECGIEKHQSSIGFSVHQLKESID 691  
QY 425 ROMIATTPSALMNGSGPETOS-----ALRGILHE-----A 454  
DB 632 RCLGLUSEQLNNYSEITNSETLVRCSHLVGLVGCYCVGVIAEEAYKSELFOKANS 751  
QY 455 AMQYKQGISFYTCRPHDIFLVARIEKVLQSGITHCABEPMKSDSKYA-----Q 505  
DB 752 LMQCGESITTLKONKTEEFRIQSLRMMQ-LCTRCLSNCTKKS--PNKIASGFFLRLTS 809  
QY 506 KYLNAKQACQRLQGY-RMPF-AMARITLFKDASGNLDKNAHPAISAIYRDSNKLSD-- 560  
DB 810 KLMNDIADICKSLASFIKKPFDGVEVSEMEDTNGNL-----MEVEDQSSMNLFFNDYD 863  
QY 561 -----DMLEKL-----ADPRKPK 574  
DB 864 SSVSDANEBGESQSTIGALNPLAEBYLSKODLLFLDMLKFLCLCVTTAQTNTVSPRAADI 923  
QY 575 MAKLPVILGNLDITIDNVSDFPNYVNSSYIPTKQETCSKTPRPE-----VEEP 625  
DB 924 RKRLMLID-----STLEPTKSLHLMYMLKELRGEYPLMEV 966  
QY 626 VBCIPKHTQPYTY-----TNHLYVPRYL--KYDSQKS----- 657  
DB 967 LELLKPLSNVCSLYRRDOVCKTILNHVLHVKNLQSNMDSBENTRDAQOGLTVIAGFM 1026  
QY 658 -----FAKARINAIQIE-FKQSDDEDSQPLKITYR--PGGVFTRSAPAVLH 705  
DB 1027 HLTGERKYIFSVMLVNLCTLLLEADPYSKMALNMWKGDFVNEVFTQ--FLADNHQ 1084  
QY 706 -----ONPEFY 711  
DB 1085 VMLMAESTINRLPOTKGDSSRLKRLPKLOQTAFENAYLKAQBSREKMSHSAENPETL 1144  
QY 712 DEIKLEPTOLHEKHLLLTFPHV--SCNNSKSGTKKRDVETQVGYMLPLKXGRV 769  
DB 1145 DBI-----YNRKSVLLTLIAVVLSCS-----PICBKQALPALKSVKENG- 1185  
QY 770 TSEOHIPVSA-NLPSGYLGYOELG--MGRHYG--PEIKWIDGKFKPLKISTH-LVSVITYQ 824  
DB 1186 --BPHLVKYLEKVSFTFYRLLEDPMASHLDYLVLEWMLNLOTDEYNLSFPFILLANTY- 1242  
QY 825 DOHLNFPYCOCT-----ESGAOLGNELVKYLKSLHA--MEGWMTAFPLTI 873  
DB 1243 --NIEDFYNSCYKVLPHLVISHRDEYVSIANQIOEDMKSLTDCFPKILVNLDPYRY 1300  
QY 874 QLFVLTTRATOREVAV-----NVRVLIHVVAQCHEEGLES-- 909

DB 1301 BGRDSGNAQORETATKYDMLKSENLGKOIDHLPISNLBEIVVELMTLHEPANSSAS 1360  
QY 910 -----HLRSYKAYKABPYVASSEKTVHEBELTSMITLKPSADF 950  
DB 1361 OSTDLCDSSGDLDPANPPHPSHVITKTPA--YISNCHKT-----KLSILEILKSPD- 1413  
QY 951 LTSNKLKYSWFFPDVLKISMAOHLIENSKYKLJRNORPPASYNHVAETVMMLPHTQ 1010  
DB 1414 -----SYCKILALCEQALETNNV--YKGRHILKITYH-----LPVSLILKDI-- 1453  
QY 1011 KFRDNPEASKANNSLAFIRKCFPTMDRGVFKOINNYISCPAGDEKTLPEYFEBL 1070  
DB 1454 -----KSGLGAMAFVLRDVIYT--LIHYNORPCIMDVSLRS-FSLCCDLIS 1499  
QY 1071 VYCNHEHYPLNLPMPFGKRIQRYQDLODYSLDBECRNHPFLVGLLRBTALOEPR 1130  
DB 1500 QVC-----OTATYCKDLLENHL-----HVIYGLI-----PLYEHOV 1532  
QY 1131 EYRLIAISVNLKNLKISFPDDRYSRSHQARIATLYLPGLILENVCRIWVDSPEPV 1190  
DB 1533 EYQKQVLDLKLKYLVIDNKDNE-----NLV-----ITIKLDPFP- 1566  
QY 1191 NAGMTVQDESLALPANDPLVTPQKSTLDNSLHKDLGAISGIASPTTSTPNINSVNA 1250  
DB 1567 -----DHVVFQDL-----RITQOKIK 1582  
QY 1251 DSRGSLITDSGNSLPERNSEKSNLGHQOQSSITGNSVNRCDKLDQSBKSL--LMCFL 1308  
DB 1583 YSRGPF-----SLBEBTNHFLVSV-----YDALPRLRLEBGLRRLQ 1621  
QY 1309 YILKMSDDELFTYNNKASTSELMDFITISEVCLHQPOYMGKRYIARTG-----MMHAR 1362  
DB 1622 BLHKQWMDIM-----RASQONPODGINVKLV--NLQLSMAINHTGEKVEAVASC 1674  
QY 1363 LQOQSLD--NSLTNHSYGHSDADYVLOSLEBANIA--TEVCLTAL-DTL-----SL 1410  
DB 1675 LGVEVPIPFSTIALQSHSDASTTKAL--KLEDEKELQWTFIMLTLYANNTLVEDCVKVS 1732  
QY 1411 FTIAFKNOLLADHGN--PLMKXVFDVYLCLFQKQKQSTA-----LKNVFTALSL 1459  
DB 1723 AVTCKNLATKYGHSFWEIYKMTTDPMLAYQPRISRKKFLVBRDKENPFEGLDI 1792  
QY 1460 -----YKPFSTFVEGRADWCAALCYEILK-CGNSKLSIRTEASQULYFLM 1505  
DB 1793 NLMITLSENHDIWITLICALFDSGTCCEL--QLKPMCEVKTDPFGQVLPYLIHDL 1850  
QY 1506 RNNFDYTKSKSFVRTHLOVISVSQLADVVGIGSTRFOQSLIINNCAUSDRLIKHTSF 1565  
DB 1851 LQDTNBSW--RMLSTHYOGFPT-----SCURHFSQTSRSTTPANL 1889  
QY 1566 SSDVD-----LTRKIRVLMATAQMKENHENDREMLVDLOYSLAKSYASTPRLKTYLD 1619  
DB 1890 DSESHFPRCCIDKKSQRTMALVUDYMRQRPS-----SGTIFNDAPWIDL 1936  
QY 1620 ---SMARITHVXNGDSEBAMCVNHTALV-AEYLTRKEAVOWEPRLPHSHSACLRBSRG 1675  
DB 1937 NYLBEVAKY-----AQSCAHTFALLYAEIYADKSSMDQ-----EKRS 1974  
QY 1727 KAERYELADIYKLIPIYERKRDPERLAHLYDTLHRAYSKVTVMHSGRL----- 1778  
DB 2025 -----KMLQPI--TRLRTEYHEA--WMGKALVTVYDELTALPSTTRQAGIIOALON 2070  
QY 1779 LGTYRVAFPQOAOYQFTDSETV-----GFFEDBDGKRYIYKPEK 1821  
DB 2071 LGLCHILSVLYKGDYENKDCPBELEHLYQAARNMOWDICTSVSKVEBGS--YHESL 2128  
QY 1822 LTPLEISQRLIKLYSDKFGSENVMMIDSGKVNPKDLDSRYAYI----- 1866

Db 2129 YNALGRLDRBFSTFYSLSKYARKEVEEMCK---SLESVSVLYPTLSRLQALGELBET 2185  
 Qy 1867 -----QTHVIFPFDEKELOERKTEPERSHNIR---FMFEMPFTQTKRGQVGEQCK 1917  
 Db 2186 GELFSFSVTH-----RQSEVYIKQKHSQSLKSDSFSQEPJAL----- 2226  
 Qy 1918 RTILTAIHCFPYKKRIPVMQHTDNLPIEVALIDEMSKVAELRQCSAEVDMIK-- 1975  
 Db 2227 -RTVI-----LEIMEKEMDNSQRECIKIDILTKHLVELSLARTFKNTQLPBR 2273  
 Qy 1976 --LQLLQGSVSVQVNAAGPLAVARAF-----LDDNTKRYPNKKK 2014  
 Db 2274 AIFQIQVNSVSGVSEWOLEEQAQVMAKESQSLALSIKQMIKKLDASCAANPS---- 2329  
 Qy 2015 LKKEVROFVACGQALA-----VNERLIKEDOLEYOEMKANY-REMAKEL 2060  
 Db 2330 -LKTYTECLRYGCMWLAEETCLENPAINVQTYIEKAVEVAGNVDGESSDEL 2379

RESULT 7.  
 US-08-642-274D-3  
 / Sequence 3. Application US/08642274D  
 / Patent No. 6200749  
 / GENERAL INFORMATION:  
 / APPLICANT: Shiloh, Yosef  
 / TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO  
 / FILE REFERENCE: 22900003  
 / CURRENT APPLICATION NUMBER: US/08/642,274D  
 / CURRENT FILING DATE: 1996-05-03  
 / NUMBER OF SEQ ID NOS: 220  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 3  
 / LENGTH: 3056  
 / TYPE: prt  
 / ORGANISM: Homo sapiens  
 US-08-642-274D-3

Query Match 1.6%; Score 177; DB 4; Length 3056;  
 Best Local Similarity 17.3%; Pred. No. 76-05;  
 Matches 435; Conservative 328; Mismatches 830; Indels 918; Gaps 118;

Qy 66 LNDCLREMLLPYDFOFOTAILRRQGYICSTVPAKAEBAQSLFTVECTIKYNS-DWHLV 124  
 Db 271 LNDLSKEVII--ELFOLQI-----YIHPKGAQTQEGK-----AYESTKRSI 311  
 Qy 125 NKRYED-----YSGFERQLNKYVKDKLPVHYVEVDEVDKDEDAASLGS 170  
 Db 312 LYNLYDLVNEISHIGSRGYSGFRNIIVKELIEIMADICHQV-----FNEDETSLEI 366  
 Qy 171 OKGGLTKHGMLYKGNNSAISVTMRSPKRRFFHLIQDGSYKFEFLKD-LOKEPKGSIF 229  
 Db 367 SOSYTTQ-----RESSDVSVCRRKK-----IELG-----WEYIKHLQKSDPFL 409  
 Qy 230 LGFL-----YGVSPNNKV-----RRFAFLK-----MODKS 256  
 Db 410 VEWLQIATOLISKYFASLPNCELSPLMLISQLPQGRGERPTPYLRCLTEVALQDGR 469  
 Qy 257 SYLLADSEVEHEWETIINKILQINFEAMOEKRGDS-----HEDDEQCK 303  
 Db 470 SNLESSQKSDLKLM---NKIWCFIRGISSEQAQENFGLLGAIIOGSLVEVDREFFK 525  
 Qy 304 L-EGSG-----SGLDSYLPMLAKSAREAEIKLSSEGRVK- 336  
 Db 526 LFTGSGACRSCRAVCCLTALTTSIVPGAVKGIENMCEVNRS-----FSLK-BSIMK 579  
 Qy 337 -LFF-LDPPAQKLDFFSABPEV--KSPFEKFGRRILVKNDLSFNLQCCVA----- 383  
 Db 580 LRFYQLEGD---LENSTEVPRILHSNFPILVEKILV-----SLTKNCKAAMNPFQSV 631  
 Qy 384 -----ENBEGPTNVPEPFV--TLSLDPI-----KYNKISADPHVDLHNSV 424  
 Db 632 ECEHHKKEELSFSEVEELFLOTTFDKMDPLIIVRECGIEKQSSIGSVHONLESID 691

Qy 425 ROMIATSPALMNGSGPETOS-----ALRGILHE-----A 454  
 Db 632 RCLLGLSGLNNTNYSSEITNSETLVRCRLLVGLVGLCYGVGVAIEEBAAYSELPQKANS 751  
 Qy 455 AMQYKQGI FSVYTCHPDI FLVARIKYLQGSITHCAEPYKSSDSKYA-----Q 505  
 Db 752 LMQAGESIITLPKNKTNEEFRIQSLRNMQ--LCRCLSNCTKKS--PNKIASGFPLRLTS 809  
 Qy 506 KYLKNAKQACORLQGY-RMPF-AMARTLPDASGNLKNARFSAIYNQDSKXLSND--- 560  
 Db 810 KLMNDIADICSLASFIKKFPDRGEVESMEDDTGNL-----MEVEDQSSNNLNDYPD 863  
 Qy 561 -----DMKLT-----ADFRKBEK 574  
 Db 864 SSVSDANPEGESQSTIGAINPLAEYLSKODLLFDMLKFLCLCVTAQTNTVSRADI 923  
 Qy 575 MAKLEVILGNLDITIDNVSSDFPNVNSYPTKQFETCSKPTTFE-----VEEF 625  
 Db 924 RKLLMLID-----SSTLEPTKSLHLMYLMMLKELPGEERYPLMEDV 966  
 Qy 626 VPCIPKHTQPYTIY-----TNHLYVPKYV--KYDSQKS----- 657  
 Db 967 LELKPLSNVCSLYRRDQVCKTIINHVLHVVKNLQGSNMDSBENTRDAQOGLTYIGAFW 1026  
 Qy 658 -----FAKARNIACIE-FKDSDEDSQPLKCIYGR--PGSPVETRSAPFAVLHH 705  
 Db 1027 HLTFRKRTI FSVRMALVNCIKTLLEADPYSKMALINWQKPFVNEVFTQ--FLADNHQ 1084  
 Qy 706 -----ONPEFY 711  
 Db 1085 VMLMAESINRLFDDTKDSSRLKALPLKLOQTAFENAYLKAQGBEMSHSANPELT 1144  
 Qy 712 DEIKIELPTQHEKHHLLTFPHV--SCDSSKSTKRDVETQVGVSWPLDKGRV 769  
 Db 1145 DEI-----YNRSAVLTLIAVLSCS-----PICEKALFALCKSVENG- 1185  
 Qy 770 TSECHI PVSA-NLPBGYLGQELG--MGRHYG-PEIKKWDGKPLKLISTH-LVSTVYVQ 824  
 Db 1186 -EHLVKKVLEKSETGYRRLBDFMASHDYLVLEMLNODTEYNLSSPFLILANT- 1242  
 Qy 825 DOHLNFPQYCOKT-----ESGAOLGNEILVKYLSLHA-MEGHVMIAFLPTLN 873  
 Db 1243 -NIEDFYRSCYKVLIPHLVIRSHPEVKSJAMQIQEOMKSLTDCPKILVNIIPYAY 1300  
 Qy 874 QLFVLTATQEVAV-----NVRVLIHVAQCEBLES-- 909  
 Db 1301 BSTRDSMAQOREIATKYVDMKSENLLGKOIIDL.FISNLPETIVELMTLHEPANSAS 1360  
 Qy 910 -----HLRSYKVAVKABPVYASEYKTVHEELTKSMTTLKPSADF 950  
 Db 1361 QSTDLCDPSGDDPAPNPHFPHSHVYKATPA--YISNCHKT--KLKSILEILKSPD- 1413  
 Qy 951 LTSNKLKYSWFFPVLIKSMAOHLIENSKYKLYLRNQPSPSYAHAVETVYVNMPLHTQ 1010  
 Db 1414 -----SYOKILAIACEQALETNNV--YKGRHILKIYH--LPVSLKLDI-- 1453  
 Qy 1011 KFRONPEASKNANSHLAVFKRCFTFEMDRGVFOINNVISCFAPGDKPTJEYFEFLR 1070  
 Db 1454 -----KSGLGAMAFVLRDVIYV--LIHYINRPSCLINDVSIRS--PSLCCDLIS 1499  
 Qy 1071 VCNHEHYIPLNLPMPFGKRIORQYODQLDYSLTDECRNHFVLGLLRLREVTALQEFR 1130  
 Db 1500 QVC-----QIATVYCKDALENHL-----HVIYGTIL-----PLVYQV 1532  
 Qy 1131 EYRLAIAVLNKLKISFDDRYASRSHQARIATLYLPGLLLENVORINVRDVSPPV 1190  
 Db 1533 EYQKQVLDLKLKYLVIDNKNQ-----NLV-----ITIKLDPFP- 1566  
 Qy 1191 NAGMTVKBESLAPRVNPLVTPQKSTLDNSLHLDLGAISGIASPYTTSTPNNINSVNA 1250  
 Db 1567 -----DHVVPKDL-----RITQOKIK 1582

Qy	1221	DSRGSILSTBGSNLSLPENSKSNLSLDHGQSSSTLGSNVACHDKLDGSEIKSL--LMCFL	1308
Db	1583	YSRGEF-----SLSEELNHPLSVSV-----YDALPPLTRLEGKDLRRQL	1621
Qy	1309	YLKMSMDALFTYWNKAKSTSELMDPFTISEVCLHQFGYMKRYIARTG-----MMHAR	1362
Db	1622	ELMHQDMWIM-----RASQNPQDGMIVKLV--NILQLSKMALNTHGEVLEAVGSC	1674
Qy	1363	LOQUGSLD-NSLTENSHYGHSDADVLYHOSLLEANIA--TEVCLTAL-DTL-----SL	1410
Db	1675	LGEVGPIDFSTIALQHSKDAAYTKAL--KLEPDEKELQWTFMYLTJNLNTLVEDCVKRYSA	1732
Qy	1411	FTLAFKQNLADHGN--PLMKKPVDFVLYCLQHQGSTA-----LKNVFTALRSL	1459
Db	1733	AVTCLKNILATKGTGSEWEIYKMTTDEMLATYLQFPRTSRKIFLEVPREDKNPEEGLDI	1792
Qy	1460	-----LYKPFSTPEYGRADCAALCYEILK-CNSKLSIRTEASQLYFLM	1505
Db	1793	NLMITPLSNMDIMWITLTCARLDSGTCCELL--QLKPKMEVKTDECGYLYPPLIHDL	1856
Qy	1506	RNNPDYTSKKSAFVRTHLQVILSVSLADVVGIGSTRFQSHLSIINNCANSDRLIKHTSF	1565
Db	1851	LQDTNESW-RNLSTHYGQFFT-----SCLRHPSQSRSTTPANL	1889
Qy	1566	SSDVED-----LYTKRITVLMATQMKHENDPEMLVDLQSLAKSYASTELKRTWID	1619
Db	1890	DSESHFPRCCLDKKSQRTMLAVUDYMRQRKPS-----SGTIFNDAFWIDL	1936
Qy	1620	---SMARITHVKNNGDLSBAMCIVNHTALV--AEYLTRKEAVQWEPRLPHSHSACLRBSRG	1675
Db	1937	NYLBYAKV-----AQSCAHHETALLYAEIYADKSSMDDQ-----EGRS	1974
Qy	1676	GVFRGCTAFRVTTPNIDEBASMMEDVGMQVHNEDVIMHL-----LEQCADGLW	1726
Db	1975	LAFEGSGOSTTL--SSLSEKSKERTGISLQ-----DLTLEIYRSIGBPSDLYGGCG--	2024
Qy	1727	KAERBELADIYKLIIPYEKRRDPERLAHLVDTLHRAYSIVTEVMHSGRL-----	1778
Db	2025	-----KMLQPI-TRIRTYENBA--MMGKALVTYDDETALPSSSTRQAGIIQALQN	2070
Qy	1779	LGYERVAFFGQAQYOFDTSEIDVE-----GPFEDQDGEVLYKEPK	1821
Db	2071	LGLCHILSVLYKGLDYENKDKCPREBELHYQAMNMOMDICTSYKSEVEGTS--YHESL	2128
Qy	1822	LTPLEISQRLIKLYSDKRGSENNVMIDSGKVNPKDLSICXAYI-----	1866
Db	2129	YNALQSLDRPRESTYESUKARVKEVEBMCK--RSLSEYSLVPTLSRLQALGELBSI	2185
Qy	1867	-----QYTHVYLPFDEKELDERKTEFEFSNHIR--PMREMPPTQGGKQGVBEQCK	1917
Db	2186	GELFERSVYTH-----QLSESVYIKWQHGQLDSDFSPOEPIPAL-----	2236
Qy	1918	RTTITLALHPFYVKGRIPVMVOHHTDNPPIEVALIDEMSKVAELRQCSSAEVDMIK--	1975
Db	2227	RTVY-----LEILMEKMDNSQRECTIDILTKHVELSTIARTPKTQTLPER	2273
Qy	1976	--LQQLKGSVSVQVNAERPLAYARAF-----LDDNTNKRYPDNKYK	2014
Db	2274	AIFOIKQJNSVSGVSEWQLBEAQFYMAKKEGSLSLSTLKQMIKKLDLASCANNPS----	2329
Qy	2015	LKKEVFRQFVEACGQALA-----VNERLIKEDOLEYOEEMKANY--REMAKEL	2060
Db	2330	-LKLTYTCTLRVCGMMLAETCLLENPAVIMQTYLEKAVAVAGNYDESSDEL	2379

RESULT 8  
US-08-952-127-3  
; Sequence 3, Application US/089522127  
; Patent No. 6211336  
; GENERAL INFORMATION:  
; APPLICANT: Shlich, Yoacel  
; APPLICANT: Tagle, Danilo A.  
; APPLICANT: Collins, Francis S.

1 TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE  
2  
3 NUMBER OF SEQUENCES: 24  
4  
5 CORRESPONDENCE ADDRESS:  
6 ADDRESSEE: Kohn & Associates  
7 STREET: 30500 No. 621136thwestern Hwy., Suite 410  
8 City: Farmington Hills  
9 State: Michigan  
10 COUNTRY: U.S.  
11  
12 ZIP: 48334  
13  
14 COMPUTER READABLE FORM:  
15 MEDIUM TYPE: Floppy disk  
16  
17 COMPUTER: IBM PC compatible  
18 OPERATING SYSTEM: PC-DOS/MS-DOS  
19 SOFTWARE: Patentn Release #1.0, Version #1.30  
20 CURRENT APPLICATION DATA:  
21 APPLICATION NUMBER: US/08/952,127  
22  
23 FILING DATE:  
24  
25 CLASSIFICATION: 530  
26  
27 ATTORNEY/AGENT INFORMATION:  
28 NAME: Kohn, Kenneth I.  
29 REGISTRATION NUMBER: 30,995  
30 REFERENCE/DOCKET NUMBER: 2290, 00029  
31  
32 TELECOMMUNICATION INFORMATION:  
33 TELEPHONE: 810-539-5050  
34  
35 TELEFAX: 810-539-5055  
36  
37 INFORMATION FOR SEQ ID NO: 3:  
38  
39 SEQUENCE CHARACTERISTICS:  
40 LENGTH: 3056 amino acids  
41 TYPE: amino acid  
42 STRANDEDNESS: single  
43 TOPOLOGY: linear  
44  
45 MOLECULE TYPE: protein  
46  
47 ORIGINAL SOURCE:  
48 ORGANISM: Homo sapiens  
49  
50 US-08-952-127-3

Query Match	1.6*	Score 177	DB 4	Length 3056
Best Local Similarity	17.3*	Pred. No. 7e-05		
Matches 435	Conservative 328	Mismatches 830	Indels 918	Gaps 118
Qy	66	LNDICREMLPPYDDFOFALRRQGRYCSFVPAKAEEBAQSLPFTESICTKYNS-DMHIV	124	
Db	271	LNDLSKEVIT---ELFQLOI---YHHKKGAKTQKG---AYESTKMSI	311	
Qy	125	NYKED-----YSGEPROLPNKVYKLDKLPVHYEYDEHVDKQEDAAASLGS	170	
Db	312	LYNYIDLLVNEISHIGSKRGKSSGGRNIAVAKENILIMADICHQ-----FNEDTRSLBI	366	
Qy	171	QKGGITTKGMLYKGNMNSAIVTMRSFRRRFFHLLQLDSDYSKPFPLD-LQKEKGSIF	229	
Db	367	QSQSYTTQ-----RESSDSVSPCKRKK-----IELG-----MEVIDHLOKSONDDL	409	
Qy	230	LGFL-----YGVSPRNKV-----RRPAELK-----MOPKS	256	
Db	410	VPMLOIATOLISKTPASLPNCSELPMLTISQLLPQHRGERTPVYVLCITVALCODKR	469	
Qy	257	SYLLAADEVEEMWITTLINKLLQLNFEAQAQEKMGDS-----HEDDEQSK	303	
Db	470	SNLSSQKSDLLKLM---NKIWCITTFPGISSEQKQENPGLGAIIOGSLVAVREFFK	525	
Qy	304	L-EGSG-----SGLDSTYLPBLAKSAREAEIKKSESRYK-	336	
Db	526	LFTGSACRSPCAVCCLTFLATLTSIVPGAVMGIEIQQNCCVNRS-----FSLK-ESIMKW	579	
Qy	337	LFP-LRDPDAQKLDPSSABEYV--KSFEKKGKRIUYKCNDSLSPYLQCVV-----	383	
Db	580	LLFQLEGD---LNSSTEVPRILHSNPFHVLVEKLV-----SLTMKQCKAAMNFFQSPV	631	
Qy	384	-----ENEGPFTNVPEPFV--TLSLFDI-----KYNRKISADPFVNDHNSFV	424	
Db	632	ECEHHHKDKKELSFSEVDELFLQTTFDKMDPLTVIREGCIKHOSSIGFVHONLKESLD	691	
Qy	425	ROMIATTSPALMNGSGPETOS-----ALRGILHE-----A	454	

```

Db      692 RCLIGSEQLNNYSEITNSETLVRCRLLVGLGACYCYMGVIAEBEAVKSELFOKANS 751
Qy      455 AMOYPKQIGSFVYCPHDFILVARIEKVLOGSITHCAEPYKMSDSSKA-----Q 505
Db      752 LMOCAGSITLFPKNTNEFRIGSLRNMQ-LCTRLCSNCTKS-PKJASGFELRLTS 809
Qy      506 KYUKNAKQACQRLGOY-RMPF-AMAAITLFKDSAGNLDKNAFSAIYRQDSNLSND--- 560
Db      810 KLMNDIADICKSLASFKKPFDGVEBSMEDDINGNL-----MEVEDOSSMLFMDYD 863
Qy      561 -----DMLKL-----ADFRKPEX 574
Db      864 SSVSDANEPESSQSTIGAINPLAEYLSKODLFLMLKFLCLCTTAQNTVSPRAADI 923
Qy      575 MAFLPILGNLDTIDNVSSDFPNVYNSAYIPKQEPETCSKPIITE-----VEBF 625
Db      924 RRLLMLID-----STLEPYSHLHNYMLMLKELPGEYPLMEVD 966
Qy      626 VPCTPKHTOPYTY-----TNHLYVPKYL--KYDSOKS----- 657
Db      967 LELKPLSNVCSLYRRDQVCKTLNHVLAHVNLGOSNMDSNTRDAGQPLTVIGATV 1026
Qy      658 -----PAKARNIAICTE-FKSDDEDSQPLKCIYGR-PGSPVFTRSAPAVLHH 705
Db      1027 HLTFRKRYIFSVMALVNCILKTLLEADPYSKAIIIVMGKDFPVNEVFTQ--FLADNHQ 1084
Qy      706 -----ONPEY 711
Db      1085 VRLMAESINRLFODTKGDSSRLKALPLKQOTAPENALYKAOBGMREKSHSAENPETL 1144
Qy      712 DEIKELPTQHEKHLLTFPHV--SCDSSKSTKSDVETQVGYSLPLKXGRVY 769
Db      1145 DEL-----YRKSVLLTLIAVLSGS-----PICQKQALPLKCSVKENG- 1185
Qy      770 TSGQHPVSA-NLPSGLYQELG--MGRHYG-PEIKWYDGKPLKISTH-LVSTVYQ 824
Db      1186 --EPHLVKKVLEKSETEFGYRRLLEDPMASHLDYLVLEMILQTEYVLSFPFILLNYT- 1242
Qy      825 DOHLNFPYCOCT-----ESGAQALGNELVKYLKSLHA-MEGVMIAFLPTIIN 873
Db      1243 --NIEPYSYCYVLLPHLVIRSHFDEBVSIAQIOEDMKSILDTGCPKILVILLYPAY 1300
Qy      874 QULFVLTRAQEEVAV-----NVTYVLIHVVAOCHEEGLES-- 909
Db      1301 EGRSDGMAQORRTATKVDMLKSENILGQIDHLFISMLPELVALLMTIHEPANSIS 1360
Qy      910 -----HLRSYKAYKAEPYVASEYKTYHEBLTKSMITTLKPSADP 950
Db      1361 QSTDLCDPSGDDLPADNPFPFESHVIAKATFA--YISNCHKT-----KLKILLEILSKSPD- 1413
Qy      951 LPSNKLKYSWPFVDLILKSMQHLSENSKVILBNQRFPAHYHAAVENTVMMLMHQIO 1010
Db      1414 -----SYQILALICEQAABETNNV--YKGRILIKTYH-----LFSVLLMDI-- 1453
Qy      1011 KPRDNBEASKNANHSLAVFIKRCFTPMDFGVFKQJNNYISCPAPGDKPTLFEYKEFPLR 1070
Db      1454 -----KSGIGGAMAFYLRDVIYT-----LIHYINQRPSCIMDSLAS-PSLCCDILS 1499
Qy      1071 VVCNHEHYPLNLPMPFGKRIQRYODLDLDSLTDFCRNHFVGLLILBEVGTALQERR 1130
Db      1500 QVC-----QTAVYCKDALENHL-----HVIYGTLI-----PLAVEQV 1532
Qy      1131 EYRLIAISVLKULLIGHSFDDRYASRSHQARTATLPLFGLLIEVQRIINRVDSPPV 1190
Db      1533 EYQKQVLDLLKYLVINKONE-----NLV-----ITIKLDPFP- 1566
Qy      1191 NAGMTVKDESLALPAVNPVTQKSGSTLNSLHKDLIGAISGIASPYTSTPINSVRA 1250
Db      1567 -----DHVFPKL-----RIQOKK 1582
Qy      1251 DSRGSLISTDSGNSLPERNSEKNSLDKIQOOSSTIGNSVVRCDKLDQSEIKSL--LMCPL 1308

```

```

Db      1583 YSRGPF-----SLLEBINHPLVSV-----YDALPLTRLEGLKDLRRQL 1621
Qy      1309 YILKMSDQALPFTYNNKASTSELMDFPISEVCLHQPOYMGKRYAFTG-----MMHAR 1362
Db      1622 ELHKQWMDIM-----RASQDNPQDGINVKLV--NLQLSKMAINHTEKEVLEAVQSC 1674
Qy      1363 LOOLSLD--NSLTFNHSYGHSDADVLHOSLEANIA--TEVCLTAL-DTL-----SL 1410
Db      1675 LGEVGPIDPSTIAIQHSADASTKAL--KLPEDEKLOMTFIMLTLYLNTIIVEDCVKAVSA 1732
Qy      1411 FTIARQQLADHGN--PLMKKVPDYUICLQKQISETA-----LKNVFTALSL 1459
Db      1733 AVTCKNLTATKTHSFWELIYKTTDPMLAYLQPRTRSKKLEVPREDKBNPFEGLDI 1792
Qy      1460 -----YKPPSTFEGRADWALCYEILK--CNSKUSIIRTBASOLYFLM 1505
Db      1793 NLMPLSENHDIWITTLCAFLDSGTCYEL--OLLRPKGVKTDFOCTULPYLIHDL 1850
Qy      1506 RNNPDTGKKSFPVTHLOVIVSOLIADVVGIGSTRFQOISLIINNCAUSDRLIKHTSF 1565
Db      1851 LQDNESW--RNLSTHVQGF-----SCLRHSQTSRSTTPANL 1889
Qy      1566 SSDVQD-----LTKIRIVLWATAQMKHENDPEMLVDLOYSLAKSYSTBELKRTWD- 1619
Db      1890 DSESHFPRCCLDKKSORTMALVVDYMRORKPS-----SGTIFNDAFWIDL 1936
Qy      1620 ---SMARJHVKNGDSEAMCVHTALV-AEYLTRKEAVQWEPPLPHSHSACLRARG 1675
Db      1937 NYLEVAKV-----NOSCAHFTALLIYEIVADKSMDDQ-----EKRS 1974
Qy      1676 GVFRQCTAFRYITENIDEASMMEDVGMQVHFNEDVJMBL-----LEOCADGLW 1726
Db      1975 LAFEESQSTI--SSLSEKSKERTGISLQ-----DLLEIYRIGRPDSLYCGGG- 2024
Qy      1727 KAERYELADIYKLIIPYERKRDPERLAHYDTLHARYSKTYEMHSGRL----- 1778
Db      2025 -----KMLQPI--TRLRTYENBA--MWGKALVYUDLETAIPSSSTRQAGIIQALON 2070
Qy      1779 LGTYFRAVFPQQAQXOFTDSEIDVE-----GFPEEDGKEVITYKEPK 1821
Db      2071 LGLCHILSYVLKGLDYENKDMCPLEBELHYOAMRNMOHCTSVSKVEGTS--YHESL 2128
Qy      1822 LTPLEISORLKLKYSDFRGSBNVMQDSGVNPKDLSKXAYI----- 1866
Db      2129 YNALQSLNDRSESTYESLKARVKEVEBMC--RSJESYSLVPTLSRLQALGELESI 2185
Qy      1867 -----QYTHVLPFDEKELDERKTEFERSHNIR--FMEDMPTQCKKQGVVEQCK 1917
Db      2186 GELFERSVTH-----RQISEVYIKQKHSQQLKDSDFSQEPITMAL----- 2226
Qy      1918 RTTILTAHCPFYVKRIIPWVOQHNTDNPJEVAIDEMSKVYABLRQCSSAEVDMIK-- 1975
Db      2227 RTVI-----LEILMEKEMDNSQREICDIULFKLIVELIARFTKQTQDBR 2273
Qy      1976 --LQKLGSVSVQVNAGPLAYARAF-----LDDTNTKRYPDNRYK 2014
Db      2274 AIFQIKQVNSVSCGVSEWQLEBAQVFMKXGOSLALSLKQMIKKLDLASCANNBS----- 2329
Qy      2015 LKXEVRFQFVEACQALA-----VNERLIKEQOLEYQEBMKANY-RENAKEL 2060
Db      2330 -LKLTYTECLRVCGMMLAETCLENPAVIMQTYLEKAVAVAGNYDESSDEL 2379

```

```

RESULT 9
US-08-952-014C-3
; Sequence 3, Application US/08952014C
; Patent No. 6265158
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yoel
; TITLE OF INVENTION: ATPXIA-TELANGIESTAIA GENE AND ITS
; GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Kohn & Associates  
 STREET: 30500 No. 626518thwestern Hwy., Suite 410  
 City: Farmington Hills  
 STATE: Michigan  
 COUNTRY: U.S.  
 ZIP: 48334  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/952,014C  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kohn, Kenneth I.  
 REGISTRATION NUMBER: 30,995  
 REFERENCE/DOCKET NUMBER: 2290,00028  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 810-539-5050  
 TELEFAX: 810-539-5055  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3056 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 US-08-952-014C-3

Query Match 1.6%; Score 177; DB 4; Length 3056;  
 Beef Local Similarity 17.3%; Pred. No. 7e-05; Indels 918; Gaps 118;  
 Matches 435; Conservative 388; Mismatches 830;

66 LNDCLREMLPPYDFOFQAIILROGRVYISTVPAKAEENOSLFTVECTITYNS-DMLV 124  
 271 LNDLSKEVIL--ELFQLQI-----YIHHPKAKTQEKQ-----AESTTMRRI 311  
 125 NYKIED-----YSGEFQQLPNKVVKLDKLPVHYVEDEVDKQEDASLGS 170  
 312 LYNLDLVNVEISHIGSRGKSSGFNIAVKENLIELMADIGHV-----FNESTSLRI 366  
 171 QKGGITKGMVLYKGNNSAIVTMRSEFKRFFHLIQDGSYKFEELK-LQKEPQSTIF 229  
 367 SGGYTTQ-----RESSDYVPCRRK-----IELG-----WEYIKHQLQKSONDFL 409  
 230 LGFL-----YGVSPNNKV-----RRFAFLK-----MQDKS 256  
 410 VPMQIATQOLISKYRPSLNCESPLMLTISQLLPQRRGERTPYVLRCLTEVALCQDGR 469  
 257 SYLLAADSEVEWEWITLTKILQLNFEAMQEKNGDS-----HEDQSK 303  
 470 SNLESQSKDLKLW-----NKIWCITFRGISSEQKQAEFGLLGAILQGSIVEVDREPK 525  
 304 L-EGSG-----SGLSYLPFLAKSAREATIKLKSERVK- 336  
 526 LFTGACRPSCAVCCLTALTSTIVPAVKMGIEQNMCEVNRS-----FSLK-BEIMTK 579  
 337 -LPY-LDPAQKLDFFSAPREV--KSPFEKFGKRLIVKCNDSLSPNQCVA----- 383  
 580 LFPQLEGD--LENSTEVPPIIHSNFPILVLEKILV-----SLTKKCKAAMNFPQSV 631  
 384 -----ENEGPTTVNEPPV--TSLSPDI-----KYNRKISADPHVDLNFPSV 424  
 632 ECEHHKDKKEELSFSEVEELFLQTFDKMDPLTIYRECGIEKQSSIGFSVQNLKESLD 691  
 425 ROMIATTSPALNNGSGEPQOS-----ALRGILH-----A 454  
 692 KCLLGLSEOLLNNYSSEITNSSETLVNCRSLLVGLVGLQCYVMGVIAEEBAKSELFOKANS 751

455 AMQYPKQGISVTCCHPDIPLVARIEKVLQGSITHCAPPMKSSDSSTVA-----Q 505  
 752 LMQAGESITLFPKNTNEEFRIQSLRNMQ-LCTRCLSNCTYKS--PNILASGFLRLTSS 809  
 506 KYLKNKQAQCCRLQY--RMPF-AMAAKTLFQDASGNLDKNARFSAIYQDSKLSND--- 560  
 810 KLMNDIADI CYSLASFIKKPFDGRGEVSEMEDTGNL-----MEVEDQSSKNLNDPDP 863  
 561 -----DMLKL-----ADPRKPEK 574  
 864 SSVSDANRGESQSTIGAINPLAEYLSKQDLPLDMLKFLCLCTTAQTNVTSRADI 923  
 575 MAKLPVILGNLDITIDNVSDFPNYSYIPTKQFCTSKPTTFE-----VEEF 625  
 924 RKKLMLID-----SSTLEPTKSLHLMYLMKLKELPGEERYPLMEBV 966  
 626 VECIRKHPQPTIY-----TNHLYVPEKYL--KYDSQKS----- 657  
 967 IELKPLSNVCSLYRRDDVCKTILNHVLHVKNLGQSNMSENRDAGQGLTVIGAFW 1026  
 658 -----FAKARVIAICIE-FKQSDSDSGLKCIYGR--PGGPVFTBSAFAVLHH 705  
 1027 HLTERTKTFVVRNALVNCLEADPYSKAILNWKQEPVAVVEFTQ--FLADNHQ 1084  
 706 -----ONPEFY 711  
 1085 VRMLAASINRLFQDTKGDSRLKALPLKQQTAFENAVYLKAGQREMSHSAENPFL 1144  
 712 DEIKLELPQLHEKHHLLTFPHV--SCDSSKSGTKKRDVETGVCSWLEPLDQGRV 769  
 1145 DEI-----YNRKSVLLTLAVVLSGS-----PICEKQALPALCKSVKENG- 1185  
 770 TSECHIPVSA-NLPSGVLYGOELG--MGRHYG-PEIKVWDGSKPLKISTH-LVSTVYVQ 824  
 1186 --EPHLVKKVLEKSETGYRLEDFMASHLDYLVLEMNLQDTEYNUSPFFILLNT- 1242  
 825 DQHLNFFQYCOKT-----ESGAQALGNELVLYKLSIHA-MEGHVMIAPLPTILN 873  
 1243 --NIEDFPRSCYKVLPLHLVIRSHDFEVSIANQIOEDMKSLTDCPRLIVNLLPYFAY 1300  
 874 QLFVLTAQGEVAV-----NTRVLIHVAQCEBGLS-- 909  
 1301 EGTDRSGMAQRETFATKYVDMKSENLLGKOIDHLFISNLPBEIVELMTLTHEPANSSAS 1360  
 910 -----HLRSYKVAYPEVYASEKTYHEBLTSMITILKPSADF 950  
 1361 OSTDLCDPSGGLDPAHPNPPSPSHVIAKTPA--YISNGHKT--KLASILELSKSPD- 1413  
 951 LTSNLLKYSWFFPDVLIKSMAOHLIENSKVLLNORFPASVYHNAVETVVMMLPHITQ 1010  
 1414 -----SYQKILALICEQALETNNV--YKGNHILKITYH--LPVSLLLKDI-- 1453  
 1011 KFRDNPASKNANHSIAVPIKRCCTFMDRGVFKOINNYISCFAPGDKPTLFYKPEFLR 1070  
 1454 -----KSGIGAMAFVLADVIYT--LIHYNORPSCIMDVLSRS-FSLCCDLIS 1499  
 1071 VVCNHEHYIPLNLPMPKGRIRQRYQDQLODLSLDFECRNHPFLVGLLRBGTALQER 1130  
 1500 QVC-----QTAVTYCKDALENHL-----HVIYGTLL--PLYEYV 1532  
 1131 EYRLIAISVLKLLIKHSFDDRYASRSHQARIATLYPLFGILLIKRVORINRVDSPPPV 1190  
 1533 EYQKVQDLDLKTLVINDKNE-----NLY-----ITIKLDPF- 1566  
 1191 NAGMTVKDESLALPAVNPVLPQKSTLDNSLHKDLALASIGIASPYTTSTPNINSVRNA 1250  
 1567 -----DHVVFQDL-----RITQOKTK 1582  
 1251 DSRGSLISTDGSNSLPERNSEKNSLKGQSSITGNSVVRCDKLDQSEIKSL-LMCTL 1308  
 1583 YSRGPF-----SLSEBINHFLSVSV-----YDALPLTLRLBSLKQDLRRQL 1621  
 1309 YLKSMDDALFTYWNKASTSELMDFTIISEVCLHQFQMGKRYIARTG-----MMHAR 1362



```

Db      1622 ELHKDQVIM-----RASQDNPDQIMVLYV--NLTQLSKAIINHTGEKEVLEAVGSC 1674
Qy      1363 LQQLSLD--NSLTFNHSYGHSDADVLHQSLEANI--TEVCUL--DTL-----SL 1410
Db      1675 LGEVGIIDSTAIQHSKQASYSKAL--KLFEDELQWPFIMLYLNTLTVEDCVKVRSA 1732
Qy      1411 FTLLAFNQLADHGH--PLMKVFPVLYCFLOKHQSEPA-----LKNVFTALRSU 1459
Db      1733 AVTCLKNILATKTHGFWMEIYKMTTPMLAYLOPFTSRKKFLVFRPKENFBSLDDI 1792
Qy      1460 -----IYFPSTFEGRADMCALCYEILK--CNSKLSIRTEASQLLYFLM 1505
Db      1793 NLMIPLSENHDIWIKLTCALFDLSCGCKEIL--QLKPKCEVKTFCQTVLYLIDHIL 1850
Qy      1506 RNNFDYTKKSFRTLOVLIISQSLADVIGIGTRFOQSLSIINNCANSDBLIGHTSF 1565
Db      1851 LQDTNBSW--RNLSTHVOGFFT-----SCLNHFQTSSTTPPALU 1889
Qy      1566 SSDVKD-----LTKRIRTVLMATAQMKHEHNDPEMLVDLOYSLAKSVASTPELAKTWLD- 1619
Db      1890 DSESHFRCCLDKSGQRTMLAVDYMRQKPS-----SGTIFMDAFWLD 1936
Qy      1620 ---SMARIHVKGDLSEAMCYVHTALV--AEYLTRKKAQVMPRLPHSHSACLRSRG 1675
Db      1937 NYLEVAKV-----AQSCAHFTALYAEIYADKKSMDQ-----EKRS 1974
Qy      1676 GVPROCTAFRVITPNIDEASMEVDVGNQDVHFNDDVLMEL-----LEQCADGLM 1726
Db      1975 LAFESQSTTI--SSLSEKSKETGISIQ-----DLLEIYRSIGEPDSLYGCGG-- 2024
Qy      1727 KAEVYELADIYKLIPIYEKRRDPERLAHLTYLRAVSKYEVNHSGRLL----- 1778
Db      2025 -----KMDPI--TRLETYEHKA--MMKALVTYDLTAIRISSRQAGIIQALON 2070
Qy      1779 LGTYFVAFFGQAQYQFTDSETV-----GPFEDDGKEYIYKEPK 1821
Db      2071 LGICHLISYVLKGLDYENKQWCELEELHYQAAMRMQMDHCTSVSEKVGTS--YHESL 2128
Qy      1822 LRFPLHSIRLKLKLYDGKSGSENVKMIQDSKYNPRDLSKVAYI----- 1866
Db      2129 YNALQSLRDEFTFESLKYARVKEVEWC--NSLSVSYSLYPTLSRLQIGLEBI 2185
Qy      1867 -----QVTHVLPFPEKELOERKTEFERSHNIR--FMFEMPQTQGRQGVGEQCK 1917
Db      2186 GELFSFSVTH-----RQISEVYIKQKHSQSLKSDSFOERIAL----- 2226
Qy      1918 RRTILTALHCFPYVKRIPVMTQHTDLNPEVALIDEMSKVAELRQLCSSAEVDMIK-- 1975
Db      2227 -RTVI-----LEILMEKEMDNSQRECIKIDILTHLLEVELSILARTFKNQQLPER 2273
Qy      1976 --LQLLQGSVSQVNAAGPLAVARAF-----LDDNTKRYPNKVK 2014
Db      2274 AIFQIQVNSVSGSEWOLEEQVFWAKKESGLSLSILKQMIKLDASCAANPS--- 2329
Qy      2015 LKAEVROVEACGQALA-----VNERLIKEDOLEYOEEMKANY--REMAKEL 2060
Db      2330 -LKLTYTECLRCVGNMLAETCLENPAVINQTYLEKAVAVAGNVGDSSSEL 2379

```

```

; SEQ ID NO 3
; LENGTH: 3057
; TYPE: PRF
; ORGANISM: Human
US-09-360-416-3

Query Match      1.6%; Score 173.5; DB 4; Length 3057;
Best local similarity 17.4%; Pred. No. 0.00014;
Matches 436; Conservative 330; Mismatches 836; Indels 899; Gaps 119;

Qy      66 LNDCLREMLPFYDFQFALRLRQGRYICSTVPAAEBEASGLFTECKTYNS--DMHLV 124
Db      273 LNDLSKEVYI--ELFQLOI-----YIHPKGAATQORQ-----AESYTKMSI 313
Qy      125 NYKED-----YSGEFPOLPNKVVDKLPVHYVEVDEEDKEDASIGS 170
Db      314 LYNLYDLVNEISHIGSGKYSGGFRNIAVENIIELMADICHQV-----FREDTRSLFI 368
Qy      171 QKGIYTKGWLKYGNNNSAISVTMRSFKRRPFHLIQLDGSYKFFLKD--IQKEPKGSIF 229
Db      369 SGTSTTQ-----RESSDYVPCRRK-----IEIG-----WEVIDHLQKQNDPDL 411
Qy      230 LGFL-----YGVFRNNKV-----RRFAELK-----MODKS 256
Db      412 VPMLOIATQLSKYFASLPNCLEPLMLISQLPQORHGERTPYVLKTEVALCODKR 471
Qy      257 SYLLAADSEVEMEWITLNLKLOLNFPAAQEKNGDS-----HEDDEQSK 303
Db      472 SNLSSQKSDLLKLM-----NKIWCITFRGISSEQIOAENFGLLCAIIGSLVEVDREPK 527
Qy      304 L-EGSG-----SGLDYLPBLAKSAREAEIKKSESRYK- 336
Db      528 LFTGACRPSRCVACCLTALTTISVPGAVKMGIGIONCEVRS-----FSLK-ESIMKM 581
Qy      337 -LFY-LDPDAQKLDSSSAPEV--KSFEKREKRLYKCNDSFPLQCCA----- 383
Db      582 LFFYQLEGD--LENSTEVPPLIHSNFPHLVLEKILV-----SLTMKCKAMNFQGVSP 633
Qy      384 -----ENEQPTNVPEPFV--TLSLPDI-----KYNRKISADFPVDNHSV 424
Db      634 ECEHHQKKEBLSFEVEBELFLQTTFDKMDPLTYREGIEGHQSIGFSVHQNLKESID 693
Qy      425 RQMIATTSBALMGSGFETOS-----ALRGILHE-----A 454
Db      694 RCLLGISEQLNNYSSEITNSSETLVRCRLVGLVGYCYMGVIAEBEAYKSELFOKANS 753
Qy      455 AMQYRKQGISVYTCHPDILFVARIKVLQGSITHCAEPYKSSPSKYA-----Q 505
Db      754 LMQCGESITLFPKNTNEEFRIQSLRNMQ--LCTRLSNCTKKS--PNKIASGFFLRLLTS 811
Qy      506 KVLKNAKQACORLQGY--RMPF--AMAAATLFDKASGNLDKNARFSAIYQDSNKL--SND-- 560
Db      812 KLMDIADI CKSLASFIKKPPDRGEVSMEDDTNKL-----MEVEDOSSNMLNNDVDP 865
Qy      561 -----DMUKLT-----ADFRKDEK 574
Db      866 SEVSQANPEBSQSTIGAINPLAEYLSKODLLFADMKFLCLCVTAQTNTVSPRAADI 925
Qy      575 MAKLEVIIGNDITIDNVS--DFPNYVNSSYIPTQGFETCSITPTFE--VEFVFCIRPH 632
Db      926 RRLKMLIDS--STLEPKSLHLNMYLLIKELGEBEY-----PLPMEDVLELLPLPSNV 977
Qy      633 TQPY-----TITNHLVYYPK-----YUKYOSKSF 658
Db      978 CSLYRRDQVCKTILNHLVAVKVLQGSNMDSENTRDQOGFLYIGAFMHLTKERKIYF 1037
Qy      659 AKARNIACIE--FKQSDSDSGLPKCIYR--FGGPFVTRSAFAVLHHN----- 705
Db      1038 SVRMALVNCVKTLEADYSKAALINVMGKQFVNVETQ--FLADNHQVXMLAASIN 1095
Qy      706 -----ONPEFYDEIKLELPQ 721
Db      1096 RLFDQTKDPSRLKALPLKLOQAFENAYLKAQGBREMSHSAENPETLDEI----- 1148

```

Oy	722	AAEKHKLITFFHV--SCDONSKGSTKKRDVETOVGSWM,PLKDGWVTSSEOHIPSA	779
Db	1149	-YNRKSUILLTILAVLSCS-----PICEKQALFALCKSVKENGEL---EPHLVKKV	1194
Oy	780	-NLPSGLYGOELG--MGRHYG-PEIKWVDGKPLKISTH-LVSTVYTODOH,HNFPQY	834
Db	1195	LEKAVSETGVRLBEDPMASHLDYLLEWMLNODTEYVNLSPFILANTY--NIEDFYRS	1251
Oy	835	COKT-----ESGQALCENELVKLTKLH-HMEHVIMAPLPTLINDLFFVLTRAT	883
Db	1252	CYKULIPHLVIRSHFDCVKSIAANOIQEDWKSLLDCEPKKILVNLLPYRAYEGTDSGMAQ	1311
Oy	884	QEEVAN-----NTRVLIHVAQA-CHEEGLES-----	909
Db	1312	QREATKTYUDMLKSENLLGKQIDHLFISNLBEIWEMLMTLHPRANSASGSTDLCBPSS	1371
Oy	910	-----HLRSYKVAKYKAEPIYASRYKVHEELTSMTTLKESADFLTSNKLKYS	960
Db	1372	DLDPAHPHPHPSHYIKATPA--YISNCHKT---KLKSIILEILSKSPD-----	1414
Oy	961	WFFPDLVILKSAOHLIENSRYKYLXNOBPASYNHVAETVNMMLPHITQKFRONPEASK	1020
Db	1415	--STOKILALCEQAAETNNV--YKXHLIKLHY--LFSLSLKDI-----K	1455
Oy	1021	NANHSIAVFIKRCFTFMDRGFVFOKINNYISCFAPGDKTLFEYKFEFLRVVCNHEHYIP	1080
Db	1456	SGLGAMAFVLRDVIYT--LIHYINQPCIMDVLSRS-FSLCDDLSSQC-----	1503
Oy	1081	LNLPMPKRGKIQKQODQLDLYSLTDFCRRNHFVLGLLREVGRLAQEFREYRLAISVL	1140
Db	1504	-QTAATYCKDLLENHL-----HVIYGLTI---PLAVEQVAVQKQVLDL	1543
Oy	1141	KNLLIKISFPDRVYASRSHQARIATLYLPLFGLLIENQRIWVRDVSPEPVNAGMTVDES	1200
Db	1544	KYLVIDNKONE-----NLX-----IITKLDPPP-----	1567
Oy	1201	LALPAVNDPLVPQKGSTLDNSLHKDLCALISGLASPYTTSIPNINSVRNADRSGLSTD	1260
Db	1568	-----DHVVFYDCL-----RITQOKIKYKSRGPF-----	1589
Oy	1261	SGNSLPERNSEKSNSLDKQOOSTLGSNSVVRCDKLDQOSEIKSL--LMCFILYIKMSMDA	1318
Db	1590	--SLLEEINHLISVY-----YDALPLTRISGLKDLRRQQLHKKQDQVNDI	1632
Oy	1319	LFTYWNKASTSELDMPFTTISEVCLHQFOYMGKRYIARTG-----MHARLQOOLGSID-N	1371
Db	1633	M-----RASQONPOQIGIMVKLVV--NLQLSKMAINHTGEKVELEAVSGCGEVPIDFS	1685
Oy	1372	SLTRNHSYGHSDADVYHOSLSLEANIA--TEVCLTAL-DTL-----SLETLAFKQDL	1420
Db	1686	TTAIOHSDKDAVYTKAL--KLFEDEKELQMTFTMLTYLNTLVLVEDCVKRSAAVTCIKILIA	1743
Oy	1421	ADHGHN--PLMKKVFVLYLCEFLQKQSEETA-----LKNVPTALASL-----	1459
Db	1744	TKTGHSFMEIYKMTTDPMLAYLOPPRTSRKKFLFVPRVIXKNPFGGLDIDIMLTPLESNH	1803
Oy	1460	---IYKPSSTYEGRADMCALCEIELK-CNSKUSISRTASQULYLYMRNNPDYTGKK	1515
Db	1804	DIWIKTLTCAFLDSGCTKCEIL--QLTKPMCEVKXDFCQJYLPYLHIDILLQDTHESW-R	1860
Oy	1516	SFVTRHLOVITISVQGLADVVGIGSTRQOGLSIINNCANSBDRLIKHSPSDVKD----	1571
Db	1861	NLSTHVGQFT-----SCLRHSQTSRSTTPPANLDSSEBHFRC	1900
Oy	1572	-LTKRIRTLVAIAQMKENHENDPEMLVDLOYSLAKSYASTPELRKWTLD-----SMARIHV	1626
Db	1901	CLDKKSQRTMLAVUDVMYRQRKPS-----SGTIFNDAFMDLNYLEAVAKV--	1945
Oy	1627	KNQDSEAMCQVNTATV--AEYLTKRAVQWEPPLPHSHSACLRRSGRGVFPQCGTAF	1685
Db	1946	-----AOSCAHHTALLVIAETIADKKSMDQ-----EKRSIAFEGSGOST	1985

Qy	1686	RVIIPNIDIEKSMMEDVJOMOVHENEVIMEL	1736
Db	1986	TI--SSISEKSKETGTSILO-----DLBEIYTSIGEPRDSLYCCGG	2025
Qy	1737	IYKLIPIYEKRRDPERLAHLYTLHRAYSKVTEVMSGRRL	1768
Db	2026	--KMLQPI-FLRLRYEHEA--MMGKALVTYDLETAIPSTROAGIIONQNLGHILSVY	2081
Qy	1769	GOAAOYQFTDSEITVE-----GFPEDECKEYIYQEPHLPULPSEISOR	1831
Db	2082	LKGLDYENKQMCPELBEELHYOAAWRNMQMDHCTSUSKVEGTS--YHRSLYNALQSLDR	2139
Qy	1832	LLKLSDPKFSGSENKOMIDSGKVNPKDLSKAYAI	1870
Db	2140	EFTFPEYSLKARVKEVEBMCK---RSLSEVYSLYPTLSRLQALGELZSIGELPSRSVTH	2196
Qy	1871	VIFPDEKELOEKTEPERSHNIR--FMFMFPFOTGRQGVSEDCRRITLTAIHC	1927
Db	2197	---ROLSEVYIKQKQSOLKDSDFGFOEIMAL-----RTVI-----	2231
Qy	1928	FPYVKKEIPWYQHTLDPNPIEVALIDMSKVAALROLCSABYDMIK---	1963
Db	2232	---LEIIMEKEMNSQRECIKDLITKHLVELSILARFPKQTPPERAIPQIKYNS	2284
Qy	1984	VSVOVNAGPLAYARAF-----LDTNTRYPNKRVKLKEVROFV	2024
Db	2285	VSCGVSEMOLEEAQVFAKKEQSLATSLKQMIKKLDASCAANNPS-----LKUTYTECL	2339
Qy	2025	EACGQALA-----VNERLIKDEQLEYOGBEMKANY--REMAKEL	2060
Db	2340	RVCGNMIAETCTLENPAVIMQYTEKAVEVAGNYSSESSDEL	2380

```

RESULT 11
US-09-360-416-2
; Sequence 2, Application US/09360416
; Patent No. 6458536
; GENERAL INFORMATION:
; APPLICANT: Richard A. Gatti
; TITLE OF INVENTION: METHODS FOR DETECTION OF ATX1A
; TITLE OF INVENTION: TELANGICTASIA MUTATIONS
; FILE REFERENCE: 510015-222
; CURRENT APPLICATION NUMBER: US/09/360,416
; CURRENT FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3056
; TYPE: PRT
; ORGANISM: Human
US-09-360-416-2

Query Match      1.6%; Score 173; DB 4; Length 3056;
Best Local Similarity 17.4%; Pred. No. 0.00016;
Matches 436; Conservative 326; Mismatches 831; Indels 918; Gaps 118;

Dy      66 LNDLRLMLFFPYDDFOTALLRRGRYICSTVPAKAEBAOSLFVYECIKTYNG-DHVLV 124
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Dd      271 LNDLSLKEVII--ELFOLQI-----YIHPKCAKTOEKG-----AYESTKMSI 311
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy      125 NYKKED-----YSGEFRQLPNKVVLDLPLVHYVEVDEVDKEDDAASIGS 170
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Dd      312 LYNLYDLVLNEISHIGSRGKTSQGFNNIAVENLIELMADIQHCV-----FNEIDTSLSI 366
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy      171 QKGGITGHGMLYKGNMNSAISVTWRSPKRRPFHILQGDGSKYKEFZLKD-LQKEPKGSIF 229
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Dd      367 SQSSTTTQ-----RESSDYSPCKRKX-----IEIG-----WEVIKDHLQKSNDFDL 409
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy      230 LGFU-----YGVSPRNKV-----RRPAFLK-----MODKS 256
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Dd      410 VPMQIQNTQLISKRPALPNCESPLMLILSQLLPQGRHGERFTYVLRCLTEVALCDPKR 469
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy      257 SYLLADSEVMEEMWITLILKILQLNFEAAMQEKRGDS-----HEDDEQSK 303
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 470 SNLESSOKSDLLKLM-----NKIWCITFRGSISSQIAENNGGLGALIGOSLVEVDEPFMK 525  
 Qy 304 L-EGSG-----SGLDYLPBLLASABEAIKUKSEBRVK- 336  
 Db 526 LFTGSACRPSCAPVCLLTALTTISIVGAVKMGIEONMCEVNS-----FSLK-ESIMXK 579  
 Qy 337 -LFFY-LDPPAQKLDFFSSABEV--KSPFEKFGRIIVKCNULSFPNLOCCA----- 383  
 Db 580 LLFYQLEGD---LENSTEVPRILHSNFPRLVLEKILV-----SLTMKNCALANFPQSV 631  
 Qy 384 -----ENEGETTVNEPFV--TSLPDI-----KYNRKISADPHVDLHFSV 424  
 Db 632 ECEHNOKDELSFSEVEELFQTFDKMDFLTIVRECGIEKQSSIGFSVHQLKESID 691  
 Qy 425 ROMIATTSPALMGSGEPTOS-----ALRGILHE-----A 454  
 Db 692 RCLGLSBOULNYSSEITNSETLVRC SRLVGLVLCYCYMGVIAEBAVKSELPQKANS 751  
 Qy 455 AMQYPRQGI FSVYTCPPHPIFLVARIEKVLGSGITTHCAEPMKSSDSKVA-----Q 505  
 Db 752 LMQCAESITLFPKNKNEBFRIGSLANMO-LCTRCLSNCTKS-PNKIASGFLRLTSS 809  
 Qy 506 KYLKNAKQAQCORIGQY-RMPF-AMAAATL FKOASGNLDKNARFSAIYRQDSNKLSD-- 560  
 Db 810 KLMNDIADICKSLASFYKKPFDRGEVSEMEDTNGNL-----MEVEDQSSMNLFPDYDP 863  
 Qy 561 -----DMKLK-----ADFKPRK 574  
 Db 864 SSVSDANBEGESOSTGAINPLAEYLSKODLLFLDMLFKL CVTTAQTNTVSFRAADI 923  
 Qy 575 MAKPLVILGNLDTIDNVSSDFPNYNSAYIPTKQETCSKTPITE-----VEEF 625  
 Db 924 RRLTLMID-----STLEPTKSLHLMYLMLELGEBEYPLMEV 966  
 Qy 626 VPCIPKHTOPYTY-----TNHLYVYRYL--KYDSQS----- 657  
 Db 967 LELKPLSNVCSLYRRDQVCKTILNHLVHLVNLGQSNMDSBENTRDAOQFLTVIGAFM 1026  
 Qy 658 -----FAKARNIAICIE-FKQSDREDSOPLKIYGR--PGGVFPRSAARAALH 705  
 Db 1027 HLTERRKYIFSVMALVNLKTLLEADPYSKMALNVGKDFVNEVFTQ--FLANHHQ 1084  
 Qy 706 -----ONPEFY 711  
 Db 1085 VMLLAESINRLFQDTRKSSRLKALPLKQOTAFENAYIKRQEGMREMSHAEPEFL 1144  
 Qy 712 DEIKIELPTQHEKHLHLLTFPHV--SCDNSSKSGTKKQDVETQVGYSLWPLLKGRV 769  
 Db 1145 DEI-----YNRKSVLLTLIAVLVLS--PICKQALPALCKSVKENG- 1185  
 Qy 770 TSEHNI PVSA-NLPSGLVGOELG--MGRHYG--PEIKWYDGGKPLKISTH-LVSVVYQ 824  
 Db 1186 --BPHLVKYLEVYSETFGRRLLEDPMASHLDYLVLEMNLQDTEYNLSFPETLANY- 1242  
 Qy 825 DOHLNFPQYCOCT-----ESGAQALGNELVKYLKSLHA-MEGVMIALFPTIIN 873  
 Db 1243 --NIBFYRSCYVLLPHLVIRSHFDEVYSIANQIOEDMKSLTLDCEPKILVNIILPYFAY 1300  
 Qy 874 QLEFRVLTATQEEVAV-----NVTREYIIVHVAQCHEEGLES-- 909  
 Db 1301 EGRDGMGAQORETATKVDMLKSENLGKQIDHLFISMLPEIIVELMTLHEPANSSAS 1360  
 Qy 910 -----HLBSYVKAAYKAEPYVASEYKTYHELETSMITLKSAAP 950  
 Db 1361 QSTDLCDFSGDDLPAADNPFPHPSHVIKATFA--YISNCHKT--KLKILLETLSKSP- 1413  
 Qy 951 LTSNKLKYSWFFDVLKISMAOHLIENSKVKLLRQRPASVYHVAVERVMMMLPHIQ 1010  
 Db 1414 -----SYQXILAIIEQAABETNNV--YKGRILIKIYH---LFSVILLDI-- 1453  
 Qy 1011 KFRDNEASKNANHSIAVFIKRCFTFMDRGFVFKQIINNYISCFAPDPTLFEYKPEFLR 1070

Db 1454 -----KSGLGAMAFVLADVIYT-----LIHIYNORPSCIMDVLSRS-FSLCCDLLS 1499  
 Qy 1071 VYCNHEHIYPLNLPMPFQKRIQRQYODLODYSLTDFECRHHFLVGLLREBGTALQEF 1130  
 Db 1500 QVC-----QTATYCKDALQNH-----HVIYGTIL---PLVYQOV 1532  
 Qy 1131 EVRLIAISVLKULLIKHSFDDRYASRSHQARIATLYLPFGILLIENVORINVRDVSPPPV 1190  
 Db 1533 EVQKVLDDLKTLVLDNKNONE-----NY-----ITIKLDDPF- 1566  
 Qy 1191 NAGMTVKDBSLAPAVNPLVTPQKSTLDNSLHKOLLGASIGIASPYTTSTPNINSVARNA 1250  
 Db 1567 -----DHVVFQDL-----RITQOKIX 1582  
 Qy 1251 DSRGSLISTDGSNSLPERNSEKSNLDKHQSSSTLGNVNRCDKDOSEIKSL--LMCEFL 1308  
 Db 1583 YSRGPF-----SLLEINHLFISVS-----YDALPLTRLEGLKDLRROL 1621  
 Qy 1309 YILKMSDDALFTYNNKASTSELMDFITISEVCLHQFOYMGKRYIARTG-----MMHAR 1362  
 Db 1622 ELHKQWMDIM-----RASQDNPQDGIWVKLV--NLQLSMAIINHTEKEVLEAVGSC 1674  
 Qy 1363 LOQLGSLD-NSLTENHSYGHSDADVLHQSLLEBANIA--TEVCLTAL-DTL-----SL 1410  
 Db 1675 LGEVGPIDFSTIAIOHSDKDASTYKPL--KLPEDKELQWTFIMLTLYANNTLVEDCVKVRSA 1732  
 Qy 1411 PTLAFKNOULDHGNH--PLMKVFPDYVLCFLQKQOSTA-----LKNVFTALBSL 1459  
 Db 1733 AVTCKNLATITGHSFWEIYKMTTDBMLAYLOPRTIRKKEFLVPRDEKPEFGLDI 1792  
 Qy 1460 -----IYKPFSTYEGRADMCALCYEILK-CCNSKLSIRTEASQULYFLM 1505  
 Db 1793 NLMIPLSENHDIWITLPCAFLDGSGTKCEIL--QLKPMCEVKTDFOQTVLPYLIHOL 1850  
 Qy 1506 RNNEDYTKKSFVRTHLOVILISVSQLADVVGIGSTRQOOSIINNANCANDRLIKHYSF 1565  
 Db 1851 LQDTNESW-RNLSLTHVOGFPT-----SCLRHSQTSRSTTPANL 1889  
 Qy 1566 SSDVD-----LTRIRITVMAQAQKHEHNDPEMLVDLOYSIASVSTPELRKTMID- 1619  
 Db 1890 DSESHFRCLDKKQSTMLAVVDYMRQRKPS-----SGTIFNDAMFIDL 1936  
 Qy 1620 ---SMARIHVKNGLDSEAMCYVHTALV-AEYLTRKSAVOWEPRLPHSHSACLRSRG 1675  
 Db 1937 NYLEVAKV-----AQCAAHFTALLVABYADKXSMDDQ-----EKRS 1974  
 Qy 1676 GVFRQCGTAFRVITPNIDEBASNMEDVGMQDVHFMEDVLMEL-----LEQCADGLM 1726  
 Db 1975 LAFEBGSGSTTI--SLSBKSKSEETGISIQ-----DILLEIYRISIGEPDLSYCGGG- 2024  
 Qy 1727 KAERYELIADYKLIIPYEKRPDERLAHYDTLHRAYSKYTEVMHSGRL----- 1778  
 Db 2025 -----KMLQPI-TRLRTYEHBA--MMGRALVYTDLETAPSSITRQGITIQAION 2070  
 Qy 1779 LGTYRVAFFQGAQOYFTDSBTVE-----GPFEDDEKGEYIYKBP 1821  
 Db 2071 LGLCHILSVYLGKLDYENKMCPELELHYQAAMNMQMDCTSVSKVEGGS--YHSBL 2128  
 Qy 1822 LTPLEISORLLKLYSDKFGSENVAMIODSGKVPKXLDKRAYI----- 1866  
 Db 2129 YNALQSLDRDEFSTYEBSLKYARVKEVEEMCK--RSLESYVSLYPTLSRLQALIGELBSI 2185  
 Qy 1867 -----OYTHVAPFDEKELOERKTEFRSHNIR--FPMEMPOTQKGGQGVBEQCK 1917  
 Db 2186 GELFRSVYH-----RQLSBEVYIKWQXHSQLSLQDSFSPQEBPMAL- 2226  
 Qy 1918 RTTILTAIHCEPVYKRIIPWYQHTDNLPIEVAIDEMSKYVABEFLQSCSSABVDMIK-- 1975  
 Db 2227 -RTV-----LEIMEKEMDNSQRECIIDITLTKHVELSILARFKNTQLPER 2273  
 Qy 1976 --LOKLQGSVSVOVNACPLAVARAF-----LDDTNTKRYDNNKYK 2014  
 Db 2274 AIFQIKQYNSVSCGVSEWQLEBAQVFAKKEGSLALSTIKQMIKKLDASCANPSS----- 2329

Qy 2015 LKVEFRQPEVACGALA-----VNERLIKEDOLEYOENKANY-REMAKEL 2060  
 Db 2330 -LKUTYTECLARVCGNMLAETCLENPAVINQTYLEKAVEVANVDGSSDEL 2379

RESULT 12

US-08-874-266-2  
 ; Sequence 2, Application US/08874266  
 ; Patent No. 5955279

GENERAL INFORMATION:

APPLICANT: Gatti, Richard A.  
 TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE  
 NUMBER OF SEQUENCES: 33  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson and Bear  
 STREET: 620 Newport Center Drive 16th Floor  
 CITY: Newport Beach  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/874,266  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mays Veneko, Nancy  
 REGISTRATION NUMBER: 36,298  
 REFERENCE/DOCKET NUMBER: UCLA006.006A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-235-8550  
 TELEFAX: 619-235-0176  
 TELEX:

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3056 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Protein  
 FRAGMENT TYPE: Internal  
 US-08-874-266-2

Query Match 1.6%; Score 172; DB 2; Length 3056;

Best Local Similarity 17.3%; Pred. No. 0.00019;  
 Matches 435; Conservative 327; Mismatches 831; Indels 918; Gaps 118;

Qy 66 LNDCLREMLLPYDQFQALIRROGRYICSTVPAKAEAEASQSLFVTECIKTYS-DMHLV 124  
 Db 271 LNDLSKVIYI---ELFQLOI-----YIHPRGAKTQHGK-----AYESTMKRSI 311

Qy 125 NYKTED-----YSGEPRQLPNKVVKLDKLPVHYVEDEVDKQEDASLGS 170  
 Db 312 LNVLYDLVNEISHISRGKYSGGFNIAVKEMLIELADICHOV-----FNEDETRSLBI 366

Qy 171 QKSGITKIGMLKGNNSAISVTMRSPKRFPHILQJDSYKFEPLKD-LQKEPKGSIF 229  
 Db 367 SCSYTTQ-----RSSDYVPCRRK-----TELJ-----WEYIKOHLQKSONDFDL 409

Qy 230 LQFL-----YGVSPFNKV-----RRPAFLK-----MODKS 256  
 Db 410 VPMLOIATOLISKYRPSLPCNELSLPMLMISQLPQRRGGRPRPVYRLCLTEVALCQDKR 469

Qy 257 SYLLAADSEVMEWTIINKLIQNFELAMQEKRGDS-----HEDDEQSK 303  
 Db 1454 -----KSLGAGAMFVLRDVIYI-----LIHYINRPSICINDVLSRS-FSLCCDILS 1499

Db 470 SNLESSQKSDLLKLM-----NKIWCITFRGISSEJOIAENFGILGAIQGSILVEVDREBWK 525

Qy 304 L-EGSG-----SGLSYLPBLAKSARBARIKLSRSRYK- 336  
 Db 526 LFTGSACRPSCAVCCLTALTTSTIVPGAVKMGIEQNMCEVNR-----PSLK-BSIMKW 579

Qy 337 -LFF-LDPPAOCLDPSAEPV--KSFEKGRKILVKNDSLSPNLQCCVA----- 383  
 Db 580 LFFQLEGD---LENSTVPRILHSNPFHLVLEKLY-----SLMKCKKAMNPFQSV 631

Qy 384 -----ENEGPPTNVPPFV--TSLPDI-----KYNKISADPRVDLNIHSV 424  
 Db 632 ECHNQKKEELSPFEVELFLQTFDKNDFLIVRECGIEGQSSISPSYQNLKESLD 691

Qy 425 ROMIATSPALMNGSGPETS-----ALNGILHE-----A 454  
 Db 692 KCLLSEOLANNYSSETTNSLTVRCRLVGVIGCTYMGVIAEEARYSELPQKANS 751

Qy 455 AMQYPRKQIFSVTCRPHDIFVARIEKVLQGSITHCAPYMKSSDSKVA-----Q 505  
 Db 752 LMQAGESITLPRKNTNEFRIGSLRNMQ-LCTRCLSNCTKKS-PNKLASGFLRLILTS 809

Qy 506 KVLKAKAQACORLQY-RMPF-AMAPATLFDASGNLDNARFSAITRQDSNKLSD--- 560  
 Db 810 KLMDIADICKSLASFIKKPFDRGESEMEDTNGNL-----MEVBDQSSMNLFNDDPD 863

Qy 561 -----DMKLK-----ADFRKREK 574  
 Db 864 SSVSDANERGESQITGAINPLAEYLSKODLFLDMKFLCLCTTACQTNTVTSFRADI 923

Qy 575 MAKLPVILGNLITIDNVSDFPNVNSSYIPTKOFETCSKTPYE-----VEEF 625  
 Db 924 RRKLMLLD-----SSTLEPTSLHMYLMMLKELRGEERYPLPMEDV 966

Qy 626 VPCIPKHTQPTIY-----TNHLVYTPKYL--KYDSQS----- 657  
 Db 967 LELLPLSNVSLYRDQVCKTILNHLVAVKNGISNMSENTRDAQGLTVIGAFW 1026

Qy 658 -----PAKARNAICR-FKQSDSDSOPKCIYGR--PGSFVTRSAFAVLHH 705  
 Db 1027 HLTERRKTIFSVRNALVNLKTLLEADPYSKWAILNVGKOPVNEVFTQ--FLDNHQQ 1084

Qy 706 -----ONPERY 711  
 Db 1085 VRMLAASINHLFOOTKGDSSRLKALPLKIQPTAFENAYLKAQGRMESHASANPEL 1144

Qy 712 DEIKTEPTQLHEKHNLLTFPHV--SCDSSKSGSTKRDVVERGVGYSMPLKQGRV 769  
 Db 1145 DEI-----YNRKSVLLTLIAVLVSCS-----PICEKQALPALCKSVKENG 1185

Qy 770 TSECHIIPIVSA-NLPSGYIGYQELG--MGRNKG-PEIKYVDGKFLPKISTH-LVSTVY 824  
 Db 1186 --EPHLVKKVLEKVSFETGYRRLLEDPMASHLDYLVLEVLNQDTEYNLSSPFIILANT- 1242

Qy 825 DQHLNPFQYCOKT-----ESGAQALGNELVLYKSLHA-MEGHVIAIAPFTILN 873  
 Db 1243 --NIDFPRSCYKVLPHLVIRSHDEYKSIANOQGEOMKSLTLTCCPKILVNLIPYAY 1300

Qy 874 QLFVLTATQEEVAV-----NTRVYIIVVAQCHEGLES-- 909  
 Db 1301 EGRTRSGMAOORETATKYDMLKSENILGKQIDHFLISNLEBEIVVELMTLHEPANSSAS 1360

Qy 910 -----HLRSYVYKVKPEPVASVYKTVHSLTSMITLKPADF 950  
 Db 1361 QSTDLCDFSGDLDPANPNPSPSHVTKATPA--YISNCHK--KLSILIEILSKSPD- 1413

Qy 951 LTSNKLKYSWFFEDVILIKSNAOHLIENSKYKLRLNCRFPSPSYHNAVETVVMMLPHITQ 1010  
 Db 1414 -----SVQKILAIQEQAAETNNV--YKGRILIKIYH-----LPVSLILKDI-- 1453

Qy 1011 KFRDNPEASKANHSLAVIFRCFTFMDRGVFKQINNYISCFAPGDKTLPFYKFEFLR 1070  
 Db 1454 -----KSLGAGAMFVLRDVIYI-----LIHYINRPSICINDVLSRS-FSLCCDILS 1499

QY 1071 VCNHEHYIPLNPMFEGKRIQRYODLDYSLTDFCSNHEVLGILLREVGTALQEFR 1130  
1500 QVC-----QTAAYVCKDALENHL-----HYVGTLL-----PLVYEQV 1532  
QY 1131 EVRLIAISVKNLILKHSFDDRYASRSHOARIATLYPLFGLLIENVORINRVDSPFPV 1190  
1533 EVQKQVLDLKYLVINKDNE-----NLV-----ITIKLDPFP- 1566  
QY 1191 MAGMTKDESLALPAPNPLVTPQKSTLDNSLHKDLGALISGIASPYTTSTPINVRNA 1250  
1567 -----BHVFKD-----RTQOKIK 1582  
QY 1251 DSRGSLISTDGSNLSERNSEKSNLSIKQOOSTLGNVVRCDKLDQSEIKSL--LMCP- 1308  
1583 YSNQPF-----SLBEINHFLSVS-----YALPLTRLEGLKDLRRD 1621  
QY 1309 YLKSMSDALFTYMNKASTSEIMDFPTSEVCLHQFYWKRYIARTG-----MMAR 1362  
1622 ELHKDQMVIM-----RASQDNPDQDIMVLYV--NLTQLSKMAINHTEKEVLEAVGSC 1674  
QY 1363 LQQLGSLD--NSLTFNHSYGHSDADVHOSLEANI--TEVCLTAL-DTL-----SL 1410  
1675 LGEVGPIDPSTAIQHSKQASYSYKAL--KLFEDEKIQWTFIMLYLNTLVEDCVKVRSA 1732  
QY 1411 FTLAFKQNLADHGHN--PLMKKVFVLYLCLQKHQSETA-----LKNVFTALRSL 1459  
1733 AVTCLKNILATKTHGHSFWEIYKWTTPMLAYLOPFRTRSKKPLEVRPFKXENFEBLDDI 1792  
QY 1460 -----IYRPSTFYEGRADMCALCYELIK--CNSKSSINTESAQLIYFLM 1505  
1793 NLMIPLSENHDIWIKTLCAFLDSGGTKCEIL--QLKPMCEVKTPFCOTVPIYLLHIDL 1850  
QY 1506 RNNFDYTGKKSFPRTLQVLIISVQLADVIGIGRRFOOSLSIINNCAINSORLIGHTSF 1565  
1851 LQDTNSW--RNLSTHVOGFT-----SCLHFSQTSSTTPPAL 1889  
QY 1566 SSDVKD-----LTKRITVLMATAQMKHEHNDPEMLVDQYSLAKSYASTPELAKTMD- 1619  
1890 DSESEHFPCCLDKQSRMTLAVDYMRQRPS-----SGTIFNDATWDL 1936  
QY 1620 ---SMARIHVKNQDLSBAMCYVHTALV--AEYLTKKAVOMERPLLPHSHSACLRSSG 1675  
1937 NYLEVAKV-----AOSCAHFTALLYAEIYADKXSMDDQ-----EKS 1974  
QY 1676 GVPRQCTAFRVTTPNIDEASWMEVDVGMQDVHFNDDVMEI-----LEQCADGLM 1726  
1975 LAFBESQSTTI--SSLSEKSKETGISIQ-----DLLEIYRSIGEPDLSYCGCG- 2024  
QY 1727 KAEVYELADIYKLIPIYEKRRDFRLAHLVDTLHRAYSKYTEVHNSGRRL----- 1778  
2025 -----KMLQPI--TRLTAYEHA--MMGKALVTYDLERTAIISSTRQAGIIQALON 2070  
QY 1779 LGTFYVAFFGQAQYQFTDSETDV-----GFPEDEGKEIYKEPK 1821  
2071 LGLCHILSYVLKGLDYENKDMCELELHYOAMRMQMCHTSVSKVEGTS--YHESL 2128  
QY 1822 LVPLEISORILKLYGDKFGSENVKMIODSGKNRPDLDSKYAYI----- 1866  
2129 YNHLQSLRBRERSTFYESLKVARAVEEWC--SLSVSYSLYPTLSRLQALIGLESI 2185  
QY 1867 -----QVTHVLPFDEKELOERKTEPERSHNIR--FMFEMPTQTSKQGVGEQCK 1917  
2186 GELFSRSVTH-----RQLSEVYIKMQHSQLLKQSDSFOERPMAL----- 2226  
QY 1918 RTIITLAIHCFPVYKRIPIVMTOHTDLNPIEVALDEMSKVAELFQCLSSAEVDNIK-- 1975  
2227 -RVI-----LBIEMKENDSORBECIKDILTKHLEVELSILARTFKNQOLPBR 2273  
QY 1976 --LQTLQGSVSVQVNAGPLAVARAF-----LDNTNTRYPPNKKV 2014  
2274 AIFQIQVNSVSCGVSEWOLEAQVWAKKEQSLALSIILKOMIKKLDASCAANNPS--- 2329

QY 2015 LKKEVRFQVBCGQALA-----VNERLIKEQOLEYQOEMKANY--RENAKEL 2060  
Db 2330 -LKTLYTEGLRVCQNWLABTCLBNPAVIMQTYLEKAVAVAGYDESSDEL 2379  
RESULT 13  
US-09-134-001C-3159  
Sequence 3159, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lymn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3159  
LENGTH: 10182  
TYPE: PRN  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159  
Query Match 1.5%; Score 167; DB 4; Length 10182;  
Best Local Similarity 16.5%; Pred. No. 0.0046;  
Matches 374; Conservative 335; Mismatches 868; Indels 688; Gaps 91;  
QY 16 FTALSKSGTAELRQSVSEVRGSLVLAKEPLIE-----PLDYENVIVOKKTOIINDCL 70  
Db 2461 PSANIKNNSQITIRPKAGQ---GNTENTNPTVIOAPAHQHTLTINEIKEGQGVNTNDI 2517  
QY 71 REMILFPYDDFQTAIRROGRTY-----CSTVPA-----KAEEPAQSLFTECIKT 116  
Db 2518 NNAVQVFN--KNRVALIQGNALPTNLGSGTSHIPVVIYSDGSSBEA---TEYRT 2569  
QY 117 YNSDHLVNYKEDYSGEFPOLPNKVYLDLPHVNYVEDEVDDEDAASIG--SQKGI 175  
Db 2570 KYNKTELINAR-----RRIDEISKENTPSSIRNPFDMANASQOINTANSADQV 2621  
QY 176 TGHGVLKGNMNSAAS--VTWRSFKRRFPHLIQDGSEYFEFKDQKEPKGISIFLG 232  
Db 2622 IGTEPATQOVNSALSKYQAQAKNKNKNAKALLQNKADNSQLVRAKEQIQQSIQPAASIDG 2681  
QY 233 LYGVSPNNKVRREPAFELKMDKSSYLLAASEVEMEETITLKIQLN-----FEAM 287  
Db 2682 MTQDSTRVYKNNRQAABQALQIHANSVINNGDATSQ-----QINDAKNTVEQAO 2729  
QY 288 QEKRGDGHEDDEQKLEGGSGGLDSYVPELAKSAREAIKLSRSRYKLFYLPDPAKL 347  
Db 2730 KDYEAKSNLRADKSQLOSAVDTLNRDV--LTNDKRPASVRKYNA--ISNIRKEIDTAKA 2786  
QY 348 DFSS--ABPEVKSFEKEFGKRIIVKCNDSLFNILQCCVAAENEBGPTTNVEBFFYLSLF 403  
Db 2787 DASSLTAKMTNSVEGVRALAK-----INTVQD----- 2814  
QY 404 DIKYNRKISADPHVDLNFVSQKMATTSPLAMNGSGEPTOSALRGLIHEAMQYPKGI 463  
Db 2815 -----KYNQALALQPKENNELVQAKKRLODADVND----- 2845  
QY 464 FSVTGRPHDIFLVARIEKVLQGSITHCAEPYKSSDSKVAQKYNKAKQACORLGOYRM 523  
Db 2846 -----IPQOQHTQOTINNVDKQREPARLTSQRYVINDGDATTOETSEKS 2893  
QY 524 PFAMARTLPKQASGNL--DKNARFSAIYRQDSNKLSDNDMLKLLADPRKPEKMAKLPIV 581  
Db 2894 KVEQAMQAL--TNAKSNLADKNELOTAYNKLIENVSTG-----KKRASIQYETA 2943  
QY 582 LQNLDTITDYNSSDFPNYVNSSYIPTQFETCSKTPITFEVBEFVPCIPKHTOPYTIYN 641

Db 2944 KARIQNDANKNEARILIGNDNPVQS-----VTOANKIKAIQPKLTALINMLON 2995  
 Qy 642 HUYVYPKLYDSOKSFAKARNIAICIEFKDSDEDSOPLKCYGRGPFVFTSAFAV 701  
 Db 2996 K-----ENNTLVNAKRL-----ENAVNTDPT----- 3019  
 Qy 702 LHHQNPFR---YDEIKIELPTQJLHEKHLILTFHFVSCDSSKSGTKRDVV----- 751  
 Db 3020 --HGMTOETINNYNAKKEAQONEI--QKANNIINNMGATAQADISEKSEKVEQVLQALQNAK 3076  
 Qy 752 -----ETQVGSYSLPPLKQGRVVTSEQHIFVSNLPSYJIGYELMGHRYGEIKM 803  
 Db 3077 NDLRADKRELOQAYNLT-----IOMVNTNKKKPSYIONY----- 3110  
 Qy 804 VDGKPELTKISTHLVSTVY--TODQHLNPFQYCOKTESGAQ-----ALGNELYVYLS 855  
 Db 3111 -----KSARNINQNTNAGNEAHNVLENTPVNAVEDALRKINAIQREVTAKAI 3162  
 Qy 856 LHAMEGHVMIAPLPTILNQLFR---VLTRATQEEVAVNTVRIIHHVACHEGLESMLR 912  
 Db 3163 LQDKEDNSELYNAKEKLDQALNSQPSLNGMTQESINNVTTK-----RR 3205  
 Qy 913 SYVKAYKAPYV-----ASEKYTHAEILTKSM-----TTILKPSADPLT- 952  
 Db 3206 EAQNIASSADTIINNGDASIEQITENKIRVEEATNALNEAKOHLTADTITSLKTEVAKLSR 3265  
 Qy 953 ---SNKLLKYSWFFDVLKISMAQHLIE--NSKYKCLRNOEFPASVHHAVETVNNML--MP 1006  
 Db 3266 RQDTNNKKSVSAYANNTHSHLOSETTQENRANTINKI-----RSVEEVNNAHENV 3320  
 Qy 1007 HTQKFRD-----NPEASKNANHSIAVFIKRCFTFPIRQFVEKOINNY----- 1049  
 Db 3321 QLNQRLDTITNLLQPLANKESLKEARNLESKINE--TVYQDMQOQSVYENYQAKIKQA 3378  
 Qy 1050 -----ISCAPBDP-----KTLFEYKEFLRVVCHENKPIPLNL--PMPFGKRIQRYD 1097  
 Db 3379 NESSIAQTLINNGDASQEVSTEIEKLNQKSELTSINHLTVNKEPRLTAKNQLQAND 3438  
 Qy 1098 LQDLYSLTQDFCNHPLVGLLREVGTAQOEFEVRLIAISVKNLLIKHSFDDRYASRS 1157  
 Db 3439 QK---PSTB-----GHTQOSVQSYERKLOEAK--DKINSINNLANN----- 3475  
 Qy 1158 HQARIATLYLPLEGLLIENVQIRINVDVSPFPVNAQMTVKDESLAL--PAVNPVTPOK 1214  
 Db 3476 -----PDVNAIRTKVETEQIINBELTQAKQGLTVQKQPLINAKTALQ 3518  
 Qy 1215 GSTLQNSLKDGLGASIASPTT-----STPNINS-----VRANDSRGSLSTDS 1261  
 Db 3519 --SLDN-----QPSITMTETATIONNAKOKAEQVIONANK---IENA 3558  
 Qy 1262 GNSLPERNSEKNSLDKHOSSSLGN--SVNRCDKLDQSEIKSLMCFIYILKMSDDML 1319  
 Db 3559 QPSVQOVSEKSK---VEQALSELNNAKSAALRADKQELCOAYNQLL----- 3601  
 Qy 1320 PTYWNKASTSELMDFETISEVCLHQFOYMGKRYIARTGMHARLQOLGSLDLSLTFNHSY 1379  
 Db 3602 -----QPTDLNNKKKPAISITVYNQRYQOFSNELNSTKTN--- 3634  
 Qy 1380 GHSADAVLHQSLEENIATEVCLTALDTLSLFLAERKQLLDHGNPLMKKYFDVY--- 1436  
 Db 3635 --TDRLIKEQNPVSADV-----NNALNKYREVOQKL 3663  
 Qy 1437 ---LCLQKHOSBTALKNVFTALRSLLYKPSFTFEYGRADMCALCYELIKONSLSLSI 1493  
 Db 3664 NEARALLQNKEDNSALVRAKEQLQAVDOVPST--EG-----MQOQTKDYNSQQA 3714  
 Qy 1494 RTEASQLLYFLMKNPDYTGKKSFWRTHQVLIISVQLADVVGIGTGFPOOLSINN 1553  
 Db 3715 QOEISK-----AQOVIDNGDATTOQSNKKTVERALBANNA 3752  
 Qy 1554 ANSDRLIKHTSPSSDVKDLTKIRTYLMTAQKKEHENDEPMIVDQYSLASYSTPEL 1613  
 Db 3753 KTGRLADKE--ELQNAVNOULTONIDTSGKTPASIRKY--NEAKSRIQTOIDSAKNEANS--- 3807

Qy 1614 RKTWLDNARIHVKNQDLSSEAMCYVHTALVAEYLTRKEAVQWEPPLPHSHSACLRBS 1673  
 Db 3808 -----ILNDNPQVS-----QVTLA-----LNKIKAVQ-----PHLDKAIAMLOKJE 3844  
 Qy 1674 RGVFRQCTARVITPNIIDEBASMEDVG-----MODVFNEDVLMEI 1717  
 Db 3845 NNNAVLQAKQLOQIVNEVDPFOGWTDTTANNYKSKKEADELEIQAKQIINNDATEQ 3904  
 Qy 1718 L-----EQCADGMKARVELIADIKLIIPIYERKDPERLAHLVDTLHRAKYTE 1770  
 Db 3905 ITNETNRVQALINAIANKN-----DLRADSQLENAVNOUL-----TON 3943  
 Qy 1771 VMSGRRLIG--TYFVAFEGQAOYQFTDSETDVGFPEDEGKEYIYKPELTLPLSEIS 1829  
 Db 3944 VDTNKKRASIQQYQAA--RQALIEQYNNASSEAHQIENS-----PSVNEVA 3990  
 Qy 1830 QRLKLYSDKFGSENVKMIQDSGVNPKDLSKVAVIQVTHVIPFEDEKE---LOERT 1885  
 Db 3991 QALQK-----VEAVQ-----LKVNDAIHILQNKENNSALVTAKN 4024  
 Qy 1886 EFERSHNIRRMFEMPTQTKRGQGVBE--QCKRTITLALHCFYVKKRIPWMOHTD 1944  
 Db 4025 QLOQSVN-----DQPLT--TGMTQDSINNYEAKREAOQAL-----RMAEAVINNGD 4069  
 Qy 1945 LNPFEVADENSK---KYAEIRQLCSSABVPMIKQLQLQGSVSVQVNAGPLAVARAFID 2001  
 Db 4070 ATAKOIS--DESKVQOALHLNDARQOULTADTTELQTVQ-----QIAR-----RG 4114  
 Qy 2002 DTNTR--RYPDNKYLKEVFRQFVEACGOALAVNERLIKEDOLEYQEBMKANYREMAKE 2059  
 Db 4115 DTNKKPRPSINAYNAKIOSLETOITSADKNANAIVOKPIRTQVQVNRALQOVN--QLNQ 4172  
 Qy 2060 LSEIMHQICPLEKTSVLPNSIHFNALISGPISTWHTGHTSSS 2104  
 Db 4173 LTELIN--QLOPLSNDALKAAKLINENKINOT---VQDTGMOQS 4213

RESULT 14  
 US-09-308-375-2  
 ; Sequence 2, Application US/09308375  
 ; Patent No. 6300117  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genencor International, Inc.  
 ; TITLE OF INVENTION: Proteases from Gram-Positive Organisms  
 ; FILE REFERENCE: GC394-PCT  
 ; CURRENT APPLICATION NUMBER: US/09/308,375  
 ; EARLIER FILING DATE: 1999-05-14  
 ; EARLIER APPLICATION NUMBER: EP9719636.4  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FaecSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2285  
 ; TYPE: prt  
 ; ORGANISM: Bacillus subtilis  
 US-09-308-375-2

Query Match 1.5%; Score 163; DB 4; Length 2285;  
 Best Local Similarity 17.0%; Pred. No. 0.00066;  
 Matches 405; Conservative 359; Mismatches 830; Indels 792; Gaps 99;

Qy 24 GRAALRSQSVSRVNGSVLAKPKLI--EPIDYE--NVIYVOKTQO---LNDCLREMLLP 77  
 Db 351 GAISGLKEMVSOALRIDTLMTNIRRVNRP--DYKTNELQSSIDGTLNKRITDILQMT 409  
 Qy 78 -----YDFQTAIRROGRYICSTVPAKAEBEASQSLFTECIKTVN-----SDWHL 123  
 Db 410 GDFGRMGDEBSLSTLTQAQVQVSDLTPODPTNLT--TRAMNFIANDSISINDK 467  
 Qy 124 VNYKYEDYSGFERQLPNKVVL--DKLPVHYVEVDEVDKEDDAISGQKGIITHGWL 182  
 Db 468 LNEVDNNYAVTTLANSIRAGSTASTFGVELNDLIGYTTAIASTRESGIV----- 521





Db 2223 LTNLSLIPNLKKT-----EIPSKTIASGDKTINLTFTFHDKLG 2263

RESULT 15

US-08-328-254-6  
Sequence 6, Application US/08328254  
Patent No. 5710022

GENERAL INFORMATION:

APPLICANT: Zhu, Xueliang  
APPLICANT: Lee, Wen-Hwa  
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CU 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-328-254-6

Query Match 1.5%; Score 162; DB 1; Length 2482;

Best Local Similarity 17.3%; Pred. No. 0.00094;

Matches 422; Conservative 340; Mismatches 830; Indels 852; Gaps 109;

QY 49 IEPLOYENIVQKQTOIILDCREMLFFYDDPQTALIRQGRYICSTYPAKAEERASGL 108  
DB 45 VIRINLHNVLDKSVETQKLAEMLOKAEFS-----DQKHQKIEEM 89  
QY 109 PTECIKTYNBWLAVNYKYEDYGEFROLPNKVVLDTL--PVHY-YEVDEVDDEDA 165  
DB 90 -----CLKTSQ-----LTQGVDELEHQLQLSNEIMKDKCYODLHAHYSLDLKSKYA 140  
QY 166 ASLGGQSGITTHGWLKGNMNSAISVTNRSFKRRFFHLIQLDGSGY--KTEFLKDLQK 222  
DB 141 SLVTNE-----DHQSLALFDDQPMHHSFANIIIGQSGMPSESECRLEADQ 188  
QY 223 EKGSTFLGLYGVSTRNNKVRFAFELKMDKSSVTLAADSVEHEEITTLINKLQAN 282  
DB 189 SPKNSIL-----QNRVSLFSLSSQKQ---MNSDLQKQEEELVQIKGEI-EEN 234  
QY 283 PEAAMQKNGSDHEDEQSKLEGSGGLDYLPEL--AKSAREAEIKLKSSESVYGLFYL 340  
DB 235 LKABEAMHQSFLAETSQRISKLOEDTSAMQNVAAETLSLENKEKELQLIND----- 286  
QY 341 DDAQKLDPSASPEVYSFEKFKRIILYKCNDSLNLQCVAAENEGPTTINVEPFTVL 400

Db 287 -----KVTEQAEIQE-----LKKSNHLI-----EDSLKQLLSBTL 319  
QY 401 SLFDIKYNRKISADPHVLDNHSVROMATATSPALMNGSGPETASL----- 447  
Db 320 SL-----EKEMSSIISLNKREIELTQE-----NCTLKEINASLQOEKKNLQKSES 367  
QY 448 -----RGLHEAMQYPRQGIFFSVTCPPHPIPLVARIEKVLQSGITTHCAPEYMKSSD 499  
Db 368 FANYIDEREKISIELSDQKQE-----KILLQRCBET-----GNAYEDLSQ 409  
QY 500 SSKVAQKVLKNAKQAC-----ORLQGYRMPFA----- 526  
Db 410 KYKAQE--KNSKCECLNBECLSLCENKRNKEIQKEKFAKEHOEFLTKLAFABERNQL 467  
QY 527 -----MAARTLPKASGMLDNA-----RPSAIYRQDSNKL-----SNDMLKL 565  
Db 468 MLEETVQALRSEKTDNQNNSKSEAGGLKQEIIMTKQENKQAEVNDLQEBQILKV 527  
QY 566 LADFRKPRKMAKL PV-----ILGNLDITIDVSSDPFYVNSSVIPTQF 610  
Db 528 MKTHCEQNLSBPIRNSGVKERESERNQCNPKQWDLVKEISLD--SY-NAQLV---QL 581  
QY 611 ETC--SKTPIPEVSEFVPCIPKTHQPY--TITYHLYVYKYLKYSQ--KSPAKARIA 665  
Db 582 EAMLNKKEIKLQESKEKECLOHELOTRGDLETSTNL-----QDMQSOEISGLKQCE 633  
QY 666 ICIEFK-----DSDEDSQPLKCIYRGGPPFTSAPFAVLHHQNEPFPDEI--- 714  
Db 634 IDAEKVIYSGPHEISTQNDNAHLQ-----SLQTTNKLNEKICEICELIQA 680  
QY 715 -KIELPQLHEKHNLLTFPVSCDSSKSGTKKRDVETQVGYSMPLKXGRVVTSEQ 773  
Db 681 EKYELVETLN-----DSRSECTATRKMAE-EVG-----KLNEVKILNDS 721  
QY 774 ---HIPVSNLPSGLYQ-----ELQMGHYGEIKYVDCGRPLKISTHL 817  
Db 722 GLHGLVEDIPGGEFGQPNQEHPSLAPLDESSEYHLLTSLDK-----EVQNH 772  
QY 818 VST-----VYTDOH--LHNPFOYCOKTESGAQALGNELVKYLSLHMEGYMIAFLPTI 871  
Db 773 ALOEKPLSLOSEHKILND--QHCQMSK-----MSELQTYVDSLKA-ENLVLSNLFNF 824  
QY 872 LNOFLRVLRATQEEVAVNVTRVIIHVAQCHEGLESLSRYKVAKABRYAKEYT 931  
Db 825 QGDIVKEN-----QGLE----- 837  
QY 932 VHEELTKSMITLIKPSADFLTSNKLKYSWFPFDVLIKMAOHLIENSKVALLNQRPPA 991  
Db 838 --EGLVPSLSSCVPDSSSLSS--LGDSSFYALBEQTCMSSLISNDEGAVSANQ----- 888  
QY 992 SYHNAVETVV--NMAMPHTQKFRNPREASKNANSLAVFPIKCTFMDRGVFQIINNY 1049  
Db 889 ---CSVDVFCSSLOEENLTRK--ETSPAPAKGVEELSLCE-----VYQSLSEK 933  
QY 1050 ISCFAPGPKTLFEKFFELRVNCHNEHYIPLNIMPFGKRIQRYODL-----QLDYS 1103  
Db 934 L-----BEKMSQGMKYE-----IQBLQGLSSERELD-- 964  
QY 1104 LTDFCRNHFVGLLRVGTALQDFREVRLLAISVLKNLLIKSFDPRVARSHQARIA 1163  
Db 965 -----CLRQYLSENEQWQOKLTSVTL-----EMESKLAEEKQTEOL 1002  
QY 1164 TLVLPGLGLIENQRIIVRQVSPRPVNAAGTVQDESALAPVNPVLPQK-----GST 1217  
Db 1003 SLELEVARLQGLD-LSSRSLLGITDQALQGRNESCIDIKENHSTETTERPKHDVHOI 1061  
QY 1218 LDNSLHKDL-----LGAISGIASPTTSTPINSVYRNASR----- 1253  
Db 1062 CDKXQQDPLNDIEKITETGALKPTGECSSGSPPTNVEPPGEBDKTQSSSEISLSFSG 1121  
QY 1254 -GSLISTD--SGN-----SLPERNSEKNS-----LDKIQOSSLGNSVVRCD-KLDQ 1297  
Db 1122 PNALVPMDFLGNQEDIHNLQLRVETSNENRLLHVLEDRDRKVESLLENKKEIDSKLHL 1181

QY 1298 SEIKSLMGLFYI-----LKSMSDALFTYMNKASTSELMDFITISEVCLHOPQMG 1349  
1182 QEVQMLTKIEACELEKIVGELKENS-----LSEKLEYS-----CDH--QELL 1225  
QY 1350 KRIYATGM-----MHA--RLOQLSLSLSTFNHSGH-----SDADVL 1387  
1226 ORVETSEGLNSDLEMHADKSSREDIG--DIVAKVNDWKERFLDVENELSRIRSEKASIE 1283  
QY 1388 HOSL-LEAN--LATE-----VCLTALDLSLFTLAFKQQLADHGNPL 1428  
1284 HENLYLEADLEVQTEKLEKDNENKOKIYCLE--BELSVYT--SERNOL--RGEIDT 1337  
QY 1429 MKKVPDYLCFLQKQSETPALKNVFTALNSLIYKPPSTFYEGRADMCALCYEILKCN 1488  
1338 MSK-----KTTALDQSEKMEKETELES--HQSECLHIOVAEAEVEKTE 1382  
QY 1489 KLSSTIRTEASQLLYFLMRNNFDYTGKSFVTHLOVIVISQLIADVIGIGTRFOQSLS 1548  
1383 LLOTLSSDVSELK-----DKTHQEKLOSLERKSQALSTKCELENOIA 1427  
QY 1549 IINNCAUSDRLIG-----TSFSDVXOL 1572  
1428 QUNK--EKELLVKESSELOARLSESDYEKLNYSKALEALVEKGEFALRLSSTOEVRHOL 1485  
QY 1573 TKRIRTV-----LMATAQMKHENDEPEMLVD-----LOYSL-----A 1604  
1486 RRGIBKLVRIBADEKKQHLIAKLERENDSLKDKVENLERLOMSEENGELVILDA 1545  
QY 1605 KSYASTPELKTWLDMSAR-----IHKNGDISEAMCYVHT 1642  
1546 ENSKAVEETLKTQIEEMARSLKVFLDLVTLRSEKENLTQIOEKQOLSELDKLSFRK 1605  
QY 1643 ALVAEYLTREAVOWEPRLPHSHSACLRSSRGVPROGCTAPRVITPNIDEEASMEDV 1702  
1606 SLLEE--KEQAEIQ-----IKESKTAVENLQNOIKE----- 1635  
QY 1703 GMQDVHFNEDV-----LMELBQCADGLWKAE-----RYELIADYKLIPI 1744  
1636 -----LNEAVALCGDQEMKATEQSLDPRIBEHQLNRSIEKLABLEADEKKQLC-V 1688  
QY 1745 YEKRRPFERLAHLVDTLHRAYSKVTEVMSGRRLTGTPRVAFPGQAQOYOTDSETV 1804  
1689 LOOLKESEHHAIDLKRVENLERELIARTNOE-----HAALAEANSKGEVETLAKIE 1742  
QY 1805 GFPEDEDEGKEY-----IYKEPKLTPLSEISORLTLKYSKGSBNVKMIODSGKVNPKD 1858  
1743 GMTQSLRGLELDVVTIRSEKENLTNLOKEQEKISELEIINSFENILOKEQEKVOMKE 1802  
QY 1859 LDSKVAYIOVTHVIPPEDEKELQERKTEBERSHINIRFMFEMPFTQTKRQGVBEQC-- 1916  
1803 KSTAMEMLOTOL-----KELNERVAL--HN-----DQACCKA 1834  
QY 1917 KRTITITAIHCPFYK-----KRIPIWYQHTD--LNPIDVIMSKVAEL-- 1962  
1835 KEQNLSSQVECELEKRAQLOGLDEAKNNYIVLQSSVNGILOVEDGKOLEKDEKREISR 1894  
QY 1963 -----ROLGSAEVDMIKQLTKLOGSAYOVNAGPLAYARAFID 2001  
1895 LKNQIODOQOLVSKLSQVGEHQLMKEQNLLENLTVLEQKIQV-----LOSKNASLQ 1948  
QY 2002 DT-----NTKRYVPDNKVKLLKEVROFVACGOALAVNERLIKEDOL-----E 2044  
1949 DTLEVLQSSYKULENLELTKQDKMSFV-----KVNKGTAKETELQREHMAQKTAE 2002  
QY 2045 YQEMKANYREMAKELSEIMHE-----QICPLEEKTSVLPNSL 2082  
2003 LOBELSGEKRLAGELQLLLEIKSSKQDLKELTLENSEIKKSL 2046

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 49.3964 Seconds  
(without alignments)  
4965.877 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936  
Sequence: 1 MSQPLLPASAEKTRKFRAL.....ISGRTPTWVHGMTSSSVV 2107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 11641973 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Published\_Applications\_AA.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10936	100.0	2107	9 US-09-815-379-8	Sequence 8, Appl1
2	10001.5	91.5	2066	9 US-09-978-244A-6	Sequence 6, Appl1
3	9926.5	90.8	1980	9 US-09-736-968A-108	Sequence 108, Appl
4	9926.5	90.8	1980	9 US-09-736-969A-94	Sequence 94, Appl1
5	9926.5	90.8	1980	9 US-09-736-960-91	Sequence 91, Appl1
6	9886.5	90.4	2040	9 US-09-978-244A-27	Sequence 27, Appl1
7	7976	72.9	1534	9 US-09-736-968A-10	Sequence 10, Appl1
8	7976	72.9	1534	9 US-09-736-969A-10	Sequence 10, Appl1
9	7976	72.9	1534	9 US-09-736-960-10	Sequence 10, Appl1
10	6784.5	62.0	1352	9 US-09-736-968A-9	Sequence 9, Appl1
11	6784.5	62.0	1352	9 US-09-736-969A-9	Sequence 9, Appl1
12	6784.5	62.0	1352	9 US-09-736-960-9	Sequence 9, Appl1
13	6375.5	58.3	2073	9 US-09-978-244A-105	Sequence 105, Appl
14	6105.5	55.8	2008	9 US-09-736-968A-105	Sequence 105, Appl
15	6105.5	55.8	2008	9 US-09-736-969A-2	Sequence 29, Appl1
16	6105.5	55.8	2008	9 US-09-736-969A-2	Sequence 29, Appl1
17	6105.5	55.8	2008	9 US-09-736-969A-91	Sequence 91, Appl1
18	6105.5	55.8	2008	9 US-09-736-960-88	Sequence 88, Appl1
19	5651	51.7	2175	9 US-09-978-244A-4	Sequence 4, Appl1

20	5560	50.8	2180	9 US-09-736-968A-110	Sequence 110, App
21	5560	50.8	2180	9 US-09-978-244A-2	Sequence 2, Appl1
22	5560	50.8	2180	9 US-09-736-969A-96	Sequence 96, Appl1
23	5560	50.8	2180	9 US-09-736-960-93	Sequence 93, Appl1
24	3608.5	33.0	1214	9 US-09-736-968A-12	Sequence 12, Appl1
25	3608.5	33.0	1214	9 US-09-736-969A-8	Sequence 8, Appl1
26	3608.5	33.0	1214	9 US-09-736-969A-12	Sequence 12, Appl1
27	3608.5	33.0	1214	9 US-09-736-960-12	Sequence 12, Appl1
28	3406.5	31.1	1288	9 US-09-736-968A-13	Sequence 13, Appl1
29	3406.5	31.1	1288	9 US-09-736-969A-13	Sequence 13, Appl1
30	3406.5	31.1	1288	9 US-09-736-960-13	Sequence 13, Appl1
31	2888	26.4	738	9 US-09-736-968A-11	Sequence 11, Appl1
32	2888	26.4	738	9 US-09-736-969A-11	Sequence 11, Appl1
33	2888	26.4	738	9 US-09-736-960-11	Sequence 11, Appl1
34	2460	22.5	2045	9 US-09-736-968A-109	Sequence 109, App
35	2460	22.5	2045	9 US-09-736-969A-95	Sequence 95, Appl1
36	2460	22.5	2045	9 US-09-736-960-92	Sequence 92, Appl1
37	2460	21.6	2047	9 US-09-736-968A-2	Sequence 2, Appl1
38	2360	21.6	2047	9 US-09-978-244A-31	Sequence 31, Appl1
39	2345	21.4	2090	9 US-09-736-968A-107	Sequence 107, App
40	2345	21.4	2090	9 US-09-978-244A-28	Sequence 28, Appl1
41	2345	21.4	2090	9 US-09-736-969A-93	Sequence 93, Appl1
42	2345	21.4	2090	9 US-09-736-960-90	Sequence 90, Appl1
43	2325	21.3	1999	9 US-09-978-244A-8	Sequence 8, Appl1
44	2242	20.5	2032	9 US-09-978-244A-12	Sequence 12, Appl1
45	2237.5	20.5	2015	9 US-09-736-968A-106	Sequence 106, App

# ALIGNMENTS

RESULT 1					
US-09-815-379-8					
Sequence 8, Application US/09815379					
Publication No. US20030073613A1					
GENERAL INFORMATION:					
APPLICANT: RASTELLI, LUCA					
TITLE OF INVENTION: GERITSEN, MARY					
TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS					
FILE REFERENCE: 10716/35					
CURRENT APPLICATION NUMBER: US/09/815.379					
CURRENT FILING DATE: 2001-03-22					
PRIOR APPLICATION NUMBER: 60/191,134					
PRIOR FILING DATE: 2000-03-22					
NUMBER OF SEQ ID NOS: 17					
SOFTWARE: Patentin Ver. 2.1					
SEQ ID NO 8					
LENGTH: 2107					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-09-815-379-8					
Query Match					
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 2107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MSQPLLPASAEKTRKFRALSKPGNAELRQSVSVVAGSVILAKPKLIBPDYENVYQ	60		
DB	1	MSQPLLPASAEKTRKFRALSKPGNAELRQSVSVVAGSVILAKPKLIBPDYENVYQ	60		
QY	61	KKTQILNDCLEMLFPDDPOTATLRQGRIGCTVPAKAEBSQSLPTECITKTVNSD	120		
DB	61	KKTQILNDCLEMLFPDDPOTATLRQGRIGCTVPAKAEBSQSLPTECITKTVNSD	120		
QY	121	WHLVYKXEDYSGEERQLPNKVVKDKLPVHYVEEVDXDBAASLGSQGGITKGM	180		
DB	121	WHLVYKXEDYSGEERQLPNKVVKDKLPVHYVEEVDXDBAASLGSQGGITKGM	180		
QY	121	WHLVYKXEDYSGEERQLPNKVVKDKLPVHYVEEVDXDBAASLGSQGGITKGM	180		
DB	121	WHLVYKXEDYSGEERQLPNKVVKDKLPVHYVEEVDXDBAASLGSQGGITKGM	180		
QY	161	LYKGMNNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIFLGLYGVSPRN	240		
DB	161	LYKGMNNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIFLGLYGVSPRN	240		
QY	161	LYKGMNNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIFLGLYGVSPRN	240		
DB	161	LYKGMNNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIFLGLYGVSPRN	240		

241 NKVRFAFELKQDKSSYLLAADSEVEMSEWITILNKILQLENEAMOEKRNDSHEDDE 300  
 241 NKVRFAFELKQDKSSYLLAADSEVEMSEWITILNKILQLENEAMOEKRNDSHEDDE 300  
 301 OSKLEBSSGGLDYLDELAKSAREAEIKLKSESRYVLFYLDPAQCLDFSSAPBVKSPF 360  
 301 OSKLEBSSGGLDYLDELAKSAREAEIKLKSESRYVLFYLDPAQCLDFSSAPBVKSPF 360  
 361 BKRGKRLVKCNDSLNFLOCCVANESEPTTNEPFFVLISLFDIKYRNKISADPHVDN 420  
 361 BKRGKRLVKCNDSLNFLOCCVANESEPTTNEPFFVLISLFDIKYRNKISADPHVDN 420  
 421 HESVRQMIATTSBALNNGSGPETQSALRGILHEAAMQPKQGISFYTCBPDIFLVARI 480  
 421 HESVRQMIATTSBALNNGSGPETQSALRGILHEAAMQPKQGISFYTCBPDIFLVARI 480  
 481 KVLQGSITTCABEPMKSSSSKVAQKVLKNAQACORLQGYRMPFAMAARTLFDKASGML 540  
 481 KVLQGSITTCABEPMKSSSSKVAQKVLKNAQACORLQGYRMPFAMAARTLFDKASGML 540  
 541 DKRARSATYRQDSNKLSDNDMLKCLADPRKPKMAKLPVILGNLDTIDNVSDEPNVY 600  
 541 DKRARSATYRQDSNKLSDNDMLKCLADPRKPKMAKLPVILGNLDTIDNVSDEPNVY 600  
 601 NSSYITPKQFETCSKPTITFEVEEFVPCIPKHTQPYTITNMLXYVPKYLKXDSQKSPAK 660  
 601 NSSYITPKQFETCSKPTITFEVEEFVPCIPKHTQPYTITNMLXYVPKYLKXDSQKSPAK 660  
 661 ARNIAICIEFKOSDEEDSOPKCTIYRPGQPVFTSRSAFAVLHHONPEFYDEIKIELPT 720  
 661 ARNIAICIEFKOSDEEDSOPKCTIYRPGQPVFTSRSAFAVLHHONPEFYDEIKIELPT 720  
 721 OLHEKHLTLTFPHVSCDSSKSGSTKRRDVEYOVYSMLPLKQDRVTSSEQHTVSN 780  
 721 OLHEKHLTLTFPHVSCDSSKSGSTKRRDVEYOVYSMLPLKQDRVTSSEQHTVSN 780  
 781 LPSGLYGOELGNGRYGPEIKKWDGKPLKISTHVLSTVYTOOHLNFPQYCKOTES 840  
 781 LPSGLYGOELGNGRYGPEIKKWDGKPLKISTHVLSTVYTOOHLNFPQYCKOTES 840  
 841 GAOALCNELVKYIKSLHAMEGHVMIAPLTLNQLFRVLTRATOEBAVNVTRVYIHHVA 900  
 841 GAOALCNELVKYIKSLHAMEGHVMIAPLTLNQLFRVLTRATOEBAVNVTRVYIHHVA 900  
 901 OCHBEGLESHLRSEYVYAKAEPIYVASEKTYHEELTKSMTTILKXSADPLSNKLKYS 960  
 901 OCHBEGLESHLRSEYVYAKAEPIYVASEKTYHEELTKSMTTILKXSADPLSNKLKYS 960  
 961 WFFFDVLKSMQHLIENSKVLLRNQRPASVYHAAVETVVMMLMPEHITQKFRDNBEAK 1020  
 961 WFFFDVLKSMQHLIENSKVLLRNQRPASVYHAAVETVVMMLMPEHITQKFRDNBEAK 1020  
 1021 NANHSLAVFTRKCFETMDRGVFKQJNNYISCFAPDPTLFEYKEEPLRVACNHEHYIP 1080  
 1021 NANHSLAVFTRKCFETMDRGVFKQJNNYISCFAPDPTLFEYKEEPLRVACNHEHYIP 1080  
 1081 LNLPMFPGGRIGRQYDLOLDVSLTDEFCRNHFLVGLLREVTALQOEPRVLLIASVL 1140  
 1081 LNLPMFPGGRIGRQYDLOLDVSLTDEFCRNHFLVGLLREVTALQOEPRVLLIASVL 1140  
 1141 KULLIIGSFDDRYSRSHQARIATLPLFGLLIENVQRIINVADSPFPVNAGMTVKXDS 1200  
 1141 KULLIIGSFDDRYSRSHQARIATLPLFGLLIENVQRIINVADSPFPVNAGMTVKXDS 1200  
 1201 LALPAVNPVTPQKSGTLDNSLHKDLGALSGIASPYTTSTPINSVRNADSGSLISTD 1260  
 1201 LALPAVNPVTPQKSGTLDNSLHKDLGALSGIASPYTTSTPINSVRNADSGSLISTD 1260  
 1261 SGNLSLBERSEKSNLSJDKHQOQSTLGNVYRCDKLOSEIKSLMLKFLYILKSMSSDALP 1320  
 1261 SGNLSLBERSEKSNLSJDKHQOQSTLGNVYRCDKLOSEIKSLMLKFLYILKSMSSDALP 1320  
 1321 TYWNKASTSBLMDFTTISEVCLHQFYMGKRYIARTGMHARLQOLGSLDNLSTFNHSYG 1380

1321 TYWNKASTSBLMDFTTISEVCLHQFYMGKRYIARTGMHARLQOLGSLDNLSTFNHSYG 1380  
 1381 HSDADVHQSLLSEANIAATEVCLTALDITSLPTLARKNOLDHGNPMLKRYFDVYLCTL 1440  
 1381 HSDADVHQSLLSEANIAATEVCLTALDITSLPTLARKNOLDHGNPMLKRYFDVYLCTL 1440  
 1441 OXHOSETALKNVFTLRSLIYKFPSTPYGRADMCAALCYETILKCCNSKLSISIRTEASOL 1500  
 1441 OXHOSETALKNVFTLRSLIYKFPSTPYGRADMCAALCYETILKCCNSKLSISIRTEASOL 1500  
 1501 LYFLMRNNDYTGKSPVRTHLOVILISVQLIADVVGIGTRFQOGLSIINNCAUSDRLI 1560  
 1501 LYFLMRNNDYTGKSPVRTHLOVILISVQLIADVVGIGTRFQOGLSIINNCAUSDRLI 1560  
 1561 KITSFSSDVKDTKXIRITVLAATAQMKHENDPEMLVDLOYSLASVASTPELRTKWLDS 1620  
 1561 KITSFSSDVKDTKXIRITVLAATAQMKHENDPEMLVDLOYSLASVASTPELRTKWLDS 1620  
 1621 MARIHVQGDLSSEAMCYVHTALVAEYLTRKEAVQMEPPLPHSHSACLRSGGVPRQ 1680  
 1621 MARIHVQGDLSSEAMCYVHTALVAEYLTRKEAVQMEPPLPHSHSACLRSGGVPRQ 1680  
 1681 GCTAFRVITPNIDEBASNMEDVGMQDVHFNEDVLMELLBOCADGLMKAREYELIADYXL 1740  
 1681 GCTAFRVITPNIDEBASNMEDVGMQDVHFNEDVLMELLBOCADGLMKAREYELIADYXL 1740  
 1741 IIPYIEKRDPFRLHLVDTLHRAVSKYTEVHNSGRLLGTYFRVAFQQAAYOQFTDSE 1800  
 1741 IIPYIEKRDPFRLHLVDTLHRAVSKYTEVHNSGRLLGTYFRVAFQQAAYOQFTDSE 1800  
 1801 TDVEGFEEDEDEKEXIYKPKLTPLEISQRLKLYSDKFGSENVKMIODSKVNPXOLD 1860  
 1801 TDVEGFEEDEDEKEXIYKPKLTPLEISQRLKLYSDKFGSENVKMIODSKVNPXOLD 1860  
 1861 SKYAVIQTHTVPIPFDEKELQERKTEFERSHNIIRPFMEPMTQTKGQGYVEOCKRRT 1920  
 1861 SKYAVIQTHTVPIPFDEKELQERKTEFERSHNIIRPFMEPMTQTKGQGYVEOCKRRT 1920  
 1921 IITAIHCPYVYKRIIPWYQHTTDLNPIEVALIDENSKVYAEIROCSSAEVMIKLOQL 1980  
 1921 IITAIHCPYVYKRIIPWYQHTTDLNPIEVALIDENSKVYAEIROCSSAEVMIKLOQL 1980  
 1981 QGSVSQVNAAGPLAARAFDLDTNTRKYPDNKVKLLKEVFRQVFAACGALAVNERLKE 2040  
 1981 QGSVSQVNAAGPLAARAFDLDTNTRKYPDNKVKLLKEVFRQVFAACGALAVNERLKE 2040  
 2041 DQLEYQEWKANYREMAKELSEIMHEQICPLEKTSVLPNSLHIFNAISGTFSTWVHGM 2100  
 2041 DQLEYQEWKANYREMAKELSEIMHEQICPLEKTSVLPNSLHIFNAISGTFSTWVHGM 2100  
 2101 TSSSSSV 2107  
 2101 TSSSSSV 2107

RESULT 2  
 US-09-978-244A-6  
 Sequence 6, Application US/09978244A  
 Publication No. US20030103992A1  
 GENERAL INFORMATION:  
 APPLICANT: Lu, Peter S  
 APPLICANT: Garman, Jonathan D.  
 APPLICANT: Candia, IIT, Albert P.  
 APPLICANT: Arbor Vita Corporation  
 TITLE OF INVENTION: CLASP MEMBRANE PROTEINS  
 FILE REFERENCE: 020554-000161US  
 CURRENT APPLICATION NUMBER: US/09/978,244A  
 CURRENT FILING DATE: 2001-10-15  
 PRIOR APPLICATION NUMBER: US 60/310,028  
 PRIOR FILING DATE: 2001-08-03  
 PRIOR APPLICATION NUMBER: US 09/737,246  
 PRIOR FILING DATE: 2000-12-13

;; PRIOR APPLICATION NUMBER: US 09/736,969  
;; PRIOR FILING DATE: 2000-12-13  
;; PRIOR APPLICATION NUMBER: US 09/736,960  
;; PRIOR FILING DATE: 2000-12-13  
;; PRIOR APPLICATION NUMBER: US 09/736,968  
;; PRIOR FILING DATE: 2000-12-13  
;; PRIOR APPLICATION NUMBER: US 60/240,545  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: US 60/240,508  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: US 60/240,503  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: US 60/240,539  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: US 60/240,543  
;; PRIOR FILING DATE: 2000-10-13  
;; Remaining Prior Application data removed - See file wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 106  
;; SOFTWARE: Patencin Ver. 2.1  
;; SEQ ID NO 6  
;; LENGTH: 2066  
;; TYPE: PRF  
;; ORGANISM: Mus sp.  
;; FEATURE:  
;; OTHER INFORMATION: CLASP-2  
US-09-978-244A-6

Query Match 91.5%; Score 10001.5; DB %; Length 2066;

Beet Local Similarity 92.5%; Pred. No. 0;  
Matches 1930; Conservative 49; Mismatches 57; Indels 51; Gaps 8;

QY 51 PUDYENVIVOKKTQILNDCLREMLLPYDPOFOTAILRQORYICTVPAKAEESQSLFV 110  
DB 1 PUDYENVIVOKKTQILNDCLREMLLPYDPOFOTAILRQORYICTVPAKAEESQSLFV 60  
QY 111 TECITVNSDMLVNYKYEDYSGEPRQLPNKYVLDKLPVHYEVEDEVDKEDAAISLGS 170  
DB 61 TECITVNSDMLVNYKYEDYSGEPRQLPNKYVLDKLPVHYEVEDEVDKEDAAISLGS 120  
QY 171 QKGGITKKGWLYKGNNSAISTVMSFKRPFHLQLGDSYKFEELKD--LQKEKGS 228  
DB 121 QKGGITKKGWLYKGNNSAISTVMSFKRPFHLQLGDSYKFEELKD--LQKEKGS 180  
QY 229 FLGFLYGVSRNNKYRPAFELMODKSSYLLAADSEVEEMWITLTKILOLNFPAHQ 288  
DB 181 FLGFLYGVSRNNKYRPAFELMODKSSYLLAADSEVEEMWITLTKILOLNFPAHQ 239  
QY 289 EKRNGDSHEDDEQSKLEGGSGSLDSTYLPFLAKSARBAEIKLKSBSRVKLFYLDPAQKLD 348  
DB 240 EKRNGDSHEDDEQSKLEGGSGSLDSTYLPFLAKSARBAEIKLKSBSRVKLFYLDPAQKLD 299  
QY 349 PSSASPEVKSFEKRGKRIIVKCNDSFNLQCCVAVNERGPTTNVPEFVITSLPDIKN 408  
DB 300 PSSASPEVKSFEKRGKRIIVKCNDSFNLQCCVAVNERGPTTNVPEFVITSLPDIKN 359  
QY 409 RKISADFHVDLNFVSROMIATTSFALMNG--SGPETQGLRGILHEAAMOTPKOIFSV 466  
DB 360 RKISADFHVDLNFVSROMIATTSFALMNG--SGPETQGLRGILHEAAMOTPKOIFSV 414  
QY 467 TCPHPDIFLVARIKYLQGSITHCAPYKSSDSRYAQVILKNKAOQORLGOYMPRA 526  
DB 415 TCPHPDIFLVARIKYLQGSITHCAPYKSSDSRYAQVILKNKAOQORLGOYMPRA 474  
QY 527 WAARTLFFKASGNLMDKNARFSAIYRQDSNKLSDMDLKLADPRKREKAKLPVILGND 586  
DB 475 WAARTLFFKASGNLMDKNARFSAIYRQDSNKLSDMDLKLADPRKREKAKLPVILGND 534  
QY 587 ITIDVNSDFPNYVNSYIPTKOFETCSKTPITFEVEFVPCIPKHTQPTTYITNHLIY 646  
DB 535 ITIDVNSDFPNYVNSYIPTKOFETCSKTPITFEVEFVPCIPKHTQPTTYITNHLIY 594  
QY 647 PLYLKYSQKSPAKARNIAICIEFKDSDEDSQPLKCIYGRPGPVFTRSAPAAVLHHQ 706  
DB 647 PLYLKYSQKSPAKARNIAICIEFKDSDEDSQPLKCIYGRPGPVFTRSAPAAVLHHQ 706

DB 595 PLYLKYSQKSPAKARNIAICIEFKDSDEDSQPLKCIYGRPGPVFTRSAPAAVLHHQ 654  
QY 707 NPEFYDEIKIELPTQJHEKHMLLTFPHVSCNSKSGSTKRDVETQVGYMLPLKXG 766  
DB 655 NPEFYDEIKIELPTQJHEKHMLLTFPHVSCNSKSGSTKRDVETQVGYMLPLKXG 714  
QY 767 RVVTSEQHI PVSANLPSGLGYQELGMRHYGPEIKWYDGGKPLKISTHLVSTVYTDQ 826  
DB 715 RVVTSEQHI PVSANLPSGLGYQELGMRHYGPEIKWYDGGKPLKISTHLVSTVYTDQ 774  
QY 827 HLANFPOYCOKTESAQALGNELVYKLSLAMEGHVIAFLPTLNDLFRVLTATQBE 886  
DB 775 HLANFPOYCOKTESAQALGNELVYKLSLAMEGHVIAFLPTLNDLFRVLTATQBE 834  
QY 887 VAVNTRVYIHVVAOCHEGLESLSRYKAYKAPVYASRYKVHBEILTSMTTILKP 946  
DB 835 VAVNTRVYIHVVAOCHEGLESLSRYKAYKAPVYASRYKVHBEILTSMTTILKP 894  
QY 947 SADFLTSNKLKYSWFFDVLIKSMAOHLIENSKVKLRLNORFPASVYHNAVETVNMMLP 1006  
DB 895 SADFLTSNKLKYSWFFDVLIKSMAOHLIENSKVKLRLNORFPASVYHNAVETVNMMLP 954  
QY 1007 HITOKFRDNPEASKXANSIAVFIKCTFPMDRGVFPQIANNYISCFAPGDKTLPKYF 1066  
DB 955 HITOKFRDNPEASKXANSIAVFIKCTFPMDRGVFPQIANNYISCFAPGDKTLPKYF 1014  
QY 1067 EFLRVVCHHEHYIPLNMPFGKRIORYODQLDYSLTDECRNHPVLGILLREVTAL 1126  
DB 1015 EFLRVVCHHEHYIPLNMPFGKRIORYODQLDYSLTDECRNHPVLGILLREVTAL 1074  
QY 1127 QEFREVRILAIISVLNKLKHSFDDRVSASHOARIATLYLPFLGLIENVOIRINRVS 1186  
DB 1075 QEFREVRILAIISVLNKLKHSFDDRVSASHOARIATLYLPFLGLIENVOIRINRVS 1134  
QY 1187 PEPVAGMTVDESLALPAVNPVTPQKSTIDNSLHODLCAIGIASPTTSPNINS 1246  
DB 1135 PEPVAGMTVDESLALPAVNPVTPQKSTIDNSLHODLCAIGIASPTTSPNINS 1194  
QY 1247 VVNAADSRLISTDSGNSLPERNSEKNSLIDHQOOSTLGNSVNCDLKDOSEISLSLMC 1306  
DB 1195 VVNAADSRLISTDSGNSLPERNSEKNSLIDHQOOSTLGNSVNCDLKDOSEISLSLMC 1253  
QY 1307 FLYILKNSDDALFVYNNKASTSELMDFETISEVCLHFOYVNGKRYIA----- 1354  
DB 1254 FLYILKNSDDALFVYNNKASTSELMDFETISEVCLHFOYVNGKRYIA----- 1313  
QY 1355 -----RTGMMHARLQGLSLDNLSTFNHSGHSDADVHLQSLLEANIATEVCIT 1403  
DB 1314 RKSQTL PVSRRNRTGMMHARLQGLSLDNLSTFNHSGHSDADVHLQSLLEANIATEVCIT 1373  
QY 1404 ALDTLSLFTLAFKONLADHGNPLMKVFPDYLCEFLGHQSEMLAKVFTLRSLIYKF 1463  
DB 1374 ALDTLSLFTLAFKONLADHGNPLMKVFPDYLCEFLGHQSEMLAKVFTLRSLIYKF 1433  
QY 1464 PSTFEGRADMCALCYEILKCNKNSLSSIRTEASQLLYFLMRNNFDYTGKSFVYTHLO 1523  
DB 1434 PSTFEGRADMCALCYEILKCNKNSLSSIRTEASQLLYFLMRNNFDYTGKSFVYTHLO 1493  
QY 1524 VIISVQLIADVIGIGTRFOOSLSIINNCAISDRILIGHTSSSDVKLTIRIVLMAV 1583  
DB 1494 VIISVQLIADVIGIGTRFOOSLSIINNCAISDRILIGHTSSSDVKLTIRIVLMAV 1553  
QY 1584 AOMKEHNDPEMLVDLOYSLAKSVASTBELRTKMLDSMARIVHKGDSLEAMCVYHVA 1643  
DB 1554 AOMKEHNDPEMLVDLOYSLAKSVASTBELRTKMLDSMARIVHKGDSLEAMCVYHVA 1613  
QY 1644 LVAEYLTRKE--AVQWEPPLPHSHSACLRRSGVROGCTARVITPNIIDEASVME 1700  
DB 1614 LVAEYLTRKE--AVQWEPPLPHSHSACLRRSGVROGCTARVITPNIIDEASVME 1673  
QY 1701 DVGMQDVHFNEDVLMELLEOCADGLWKAERYELADIYKLIIPYEKRDPERLAHLYDT 1760  
DB 1674 DVGMQDVHFNEDVLMELLEOCADGLWKAERYELADIYKLIIPYEKRDPERLAHLYDT 1733

QY 1761 LHRVSKTEVMSHGRLLGTYPFVAFGQAAQYQFTDSETEGFEDEGKEIYKPE 1820  
DB 1734 LHRVSKTEVMSHGRLLGTYPFVAFGQ-----GFEDEGKEIYKPE 1779  
QY 1821 KLTPLSEISORLKLKYSDFKGSENVMTIDSGKVNPKDLDSKAYIQTHTVIFPEDEKEI 1880  
DB 1780 KLTPLSEISORLKLKYSDFKGSENVMTIDSGKVNPKDLDSKAYIQTHTVIFPEDEKEI 1839  
QY 1881 QERKTEFESHNRIRRMFEMFPTQTGKQGVGEQCKRTILTAIHCFPVKKRIEVMQ 1940  
DB 1840 QERKTEFESHNRIRRMFEMFPTQTGKQGVGEQCKRTILTAIHCFPVKKRIEVMQ 1899  
QY 1941 HHTDLNPIEVAIDEMSKVAELRQLSCSAEVDMIKQLKQGSVSVQVNAAPLAVARAF 2000  
DB 1900 HHTDLNPIEVAIDEMSKVAELRQLSCSAEVDMIKQLKQGSVSVQVNAAPLAVARAF 1959  
QY 2001 DDNTKRYPDNKKLKEVFRQVEACGQALANNEKLKEDQLEFQDEMKANTRENAKEL 2060  
DB 1960 DDNTKRYPDNKKLKEVFRQVEACGQALANNEKLKEDQLEFQDEMKANTRENAKEL 2019  
QY 2061 SEIMHBOICPLEKTSVLPNSLHIFNAISGTPSTVWGHWTSSSVV 2107  
DB 2020 SDIMRBOICPLEKTSVLPNSLHIFNAISGTPSTVWGHWTSSSVV 2066

RESULT 3  
US-09-736-968A-108  
Sequence 108 Application US/09736968A  
Patent No. US20020169283A1  
GENERAL INFORMATION:  
APPLICANT: Lu, Peter  
APPLICANT: Garman, Jonathan David  
APPLICANT: Candia, III, Albert Frederick  
APPLICANT: Arbor Vita Corporation  
TITLE OF INVENTION: CLASP-7 Transmembrane Protein  
FILE REFERENCE: 020054-00061105  
CURRENT APPLICATION NUMBER: US/09/736,968A  
PRIOR FILING DATE: 2000-12-13  
PRIOR APPLICATION NUMBER: US 60/160,860  
PRIOR FILING DATE: 1999-10-21  
PRIOR APPLICATION NUMBER: US 60/162,498  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 60/170,453  
PRIOR FILING DATE: 1999-12-13  
PRIOR APPLICATION NUMBER: US 60/176,195  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: US 60/182,296  
PRIOR FILING DATE: 2000-02-14  
PRIOR APPLICATION NUMBER: US 09/547,276  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,267  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,460  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,527  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,528  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 09/687,837  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,503  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,508  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,539  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,543  
NUMBER OF SEQ ID NOS: 115  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 108  
LENGTH: 1980

TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human CLASP-2  
US-09-736-968A-108

Query Match 90.8%; Score 9926.5; DB 9; Length 1980;  
Best Local Similarity 93.8%; Pred. No. 0;  
Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;

QY 73 MLFPYDDPQTAIILRRQGYICSTYPAKAEBAOSLFTECIKTYSNDMHLVNYKEDXS 132  
DB 1 MLFPYDDPQTAIILRRQGYICSTYPAKAEBAOSLFTECIKTYSNDMHLVNYKEDXS 60  
QY 133 GEFROLPNKVVLDLPLVHYVEVDEVDKEDDAALSGQKGIITGHGMLYKGNMSAISV 192  
DB 61 GEFROLPNKVVLDLPLVHYVEVDEVDKEDDAALSGQKGIITGHGMLYKGNMSAISV 120  
QY 193 TWSFRARFPHLIQLGDSYKFEPLKD--LQKPKGISIFLGLYGVSPRNNKVRPAPFL 250  
DB 121 TWSFRARFPHLIQLGDSYKFEPLKD--LQKPKGISIFLGLYGVSPRNNKVRPAPFL 179  
QY 251 KMQDKSSYLLADSEVMEEMWITTIINKILQINPEAAMQKRGDSDHEDDEQSKLGSGSG 310  
DB 180 KMQDKSSYLLADSEVMEEMWITTIINKILQINPEAAMQKRGDSDHEDDEQSKLGSGSG 239  
QY 311 LDSYLPDLAKSAREAEIKLKSSEKRVKLFYLDPAQKLDSSAEPKVSFEKFGRIILYK 370  
DB 240 LDSYLPDLAKSAREAEIKLKSSEKRVKLFYLDPAQKLDSSAEPKVSFEKFGRIILYK 299  
QY 371 CNDLSFNIQLCCVAENEBEGTTNVEPPFTVLSLFDIKYNRKISADPHVDLHPSVRQMIAT 430  
DB 300 CNDLSFNIQLCCVAENEBEGTTNVEPPFTVLSLFDIKYNRKISADPHVDLHPSVRQMIAT 359  
QY 431 TSPALMNGSGPRTQALGIIHEAAMQPKQISFVTCGHPDIFIVARIEKLGQSIITHC 490  
DB 360 TSPALMNGSG--QSPVLKGIHEAAMQPKQISFVTCGHPDIFIVARIEKLGQSIITHC 418  
QY 491 AEPYMKSSDSSKYAAKVLKNAKQACORLQGYMPMAARLTFFKASGLDNARFSAIY 550  
DB 419 AEPYMKSSDSSKYAAKVLKNAKQACORLQGYMPMAARLTFFKASGLDNARFSAIY 478  
QY 551 RODSNKLSNDMLKLLADFRKPEKNAKLPVILGNDITIDTIVNSDPPNYVNSSYIPTKOF 610  
DB 479 RODSNKLSNDMLKLLADFRKPEKNAKLPVILGNDITIDTIVNSDPPNYVNSSYIPTKOF 538  
QY 611 ETCSTPTTFEEVEEVPVPCIPKHTOPYTYTNHLYVYPKLYKXDSQKSPKANNIACIEF 670  
DB 539 ETCSTPTTFEEVEEVPVPCIPKHTOPYTYTNHLYVYPKLYKXDSQKSPKANNIACIEF 598  
QY 671 KDSDEDSQPLKCIYGRPGPVFTSAPFAVLHHHQNPEFYDEIKIELPTQLHEKHLL 730  
DB 599 KDSDEDSQPLKCIYGRPGPVFTSAPFAVLHHHQNPEFYDEIKIELPTQLHEKHLL 658  
QY 731 TFFHVSNDSSKSGSTKGDVAVETQVGYMLPLKQGRVVTSSQHIIPVSNLPSGYCYOE 790  
DB 659 TFFHVSNDSSKSGSTKGDVAVETQVGYMLPLKQGRVVTSSQHIIPVSNLPSGYCYOE 718  
QY 791 LGMGRHYGEIIMVNDGKPLKISTHILVSTVYTOQHILNFQYOKKTESSGQALGNELV 850  
DB 719 LGMGRHYGEIIMVNDGKPLKISTHILVSTVYTOQHILNFQYOKKTESSGQALGNELV 778  
QY 851 KYLKSILHMEGHVMIAPLPTINQLFRVLTTRAQSEAVANVTYRVIILHYVAQCHEGLES 910  
DB 779 KYLKSILHMEGHVMIAPLPTINQLFRVLTTRAQSEAVANVTYRVIILHYVAQCHEGLES 838  
QY 911 LRSYKYAVKABPYVASEKTVHEELTKSMTTILKPSADFLTSNKLKLSWFFPVLKIS 970  
DB 839 LRSYKYAVKABPYVASEKTVHEELTKSMTTILKPSADFLTSNKLKLSWFFPVLKIS 898  
QY 971 MAQHLIENSKVTLNQRFPASVYHAAVETVYVNMMLPHITQKGRNDPEASKNHNHSLAVPI 1030  
DB 899 MAQHLIENSKVTLNQRFPASVYHAAVETVYVNMMLPHITQKGRNDPEASKNHNHSLAVPI 958

```
Qy 1031 KRCFTMDRGVFKQJNNYISCPAGDPKTLFEYKEFLRIVCNHEHYIPLNLPMPFGKG 1090
Db 959 KRCFTMDRGVFKQJNNYISCPAGDPKTLFEYKEFLRIVCNHEHYIPLNLPMPFGKG 1018
Qy 1091 RIORVODLDVSLDEFCRNHPLVGLLRVGTALQEBRYVLLAISVLKULLIKHSPD 1150
Db 1019 RIORVODLDVSLDEFCRNHPLVGLLRVGTALQEBRYVLLAISVLKULLIKHSPD 1078
Qy 1151 DRYASHOARIATLYPLFGLLIENVORINRVDSPFPVNAGMTVKDESLPAVNPV 1210
Db 1079 DRYASHOARIATLYPLFGLLIENVORINRVDSPFPVNAGMTVKDESLPAVNPV 1138
Qy 1211 TPQKSTLDNSLHKDLGAIISGIASPYTSTPNINSVRNADSGSLISTDSGNSLPERNS 1270
Db 1139 TPQKSTLDNSLHKDLGAIISGIASPYTSTPNINSVRNADSGSLISTDSGNSLPERNS 1198
Qy 1271 EKSNSLDKQOOSTLGNSVVRCDKLDQSEIKSLMCFYLILKMSDDALFTWNKASTSE 1330
Db 1199 EKSNSLDKQOOSTLGNSVVRCDKLDQSEIKSLMCFYLILKMSDDALFTWNKASTSE 1258
Qy 1331 LMDPFTISEVCLHOFQYMGKRYA-----RTGMHARLQOLG 1367
Db 1259 LMDPFTISEVCLHOFQYMGKRYA-----RTGMHARLQOLG 1318
Qy 1368 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDTLSLFTLAFKQNLADGHP 1427
Db 1319 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDTLSLFTLAFKQNLADGHP 1378
Qy 1428 LMKKVFDDVYLCFLQKHQOSETALKNFTALRSILYKFPSTVEGRADMCALCYEILKCN 1487
Db 1379 LMKKVFDDVYLCFLQKHQOSETALKNFTALRSILYKFPSTVEGRADMCALCYEILKCN 1438
Qy 1488 SKLSSRTAESOLVLYMNNNDPYGKSFVRTHLOVIVSGLADVIGIGSTRQOGL 1547
Db 1439 SKLSSRTAESOLVLYMNNNDPYGKSFVRTHLOVIVSGLADVIGIGSTRQOGL 1498
Qy 1548 SIINNCAANDRLIKHTSPSSDVLDLTKIRTVLMATAQKCEHNDPEMLVDLOYSIAKSY 1607
Db 1499 SIINNCAANDRLIKHTSPSSDVLDLTKIRTVLMATAQKCEHNDPEMLVDLOYSIAKSY 1558
Qy 1608 ASTPELRKWTLDSMARIVYNGDLSBAMCYHVTALVABYLTRKAVOWEPPLPHSHS 1667
Db 1559 ASTPELRKWTLDSMARIVYNGDLSBAMCYHVTALVABYLTRKAVOWEPPLPHSHS 1603
Qy 1668 ACLRSRGVFPQGCAPFVITPNIDEASMMEDVGMQVHFVEDVLMELLOQCADGLMK 1727
Db 1604 -----GVFRQGCAPFVITPNIDEASMMEDVGMQVHFVEDVLMELLOQCADGLMK 1655
Qy 1728 AERYELIADIYKLIIPYEKRRDFERLALYDILHRAYSKTEVMHSGRLLGTYPFVAF 1787
Db 1656 AERYELIADIYKLIIPYEKRRD----- 1678
Qy 1788 PQOAAQYOFPTDSETVGFFPEDEGKEYIYKEPKLTPLEISIQRLIKLYSDKFGSENVKM 1847
Db 1679 -----FFPEDEGKEYIYKEPKLTPLEISIQRLIKLYSDKFGSENVKM 1720
Qy 1848 IODSGKVNPKDLDKRAYIQVTHVIPFPBEKELQEKTIPESSHNRIRRMFEMPFTQTK 1907
Db 1721 IODSGKVNPKDLDKRAYIQVTHVIPFPBEKELQEKTIPESSHNRIRRMFEMPFTQTK 1780
Qy 1908 ROGVEEOGKRRITLAIHCFPVYKRIIPVMTQHTDMLPIEVAIDEMSKVAYELBQLOS 1967
Db 1781 ROGVEEOGKRRITLAIHCFPVYKRIIPVMTQHTDMLPIEVAIDEMSKVAYELBQLOS 1840
Qy 1968 SAEVMDIKLQLKQSSVQVNAAGPLAVALADTNTGRYPDNVYKLLKEYFRQFVAC 2027
Db 1841 SAEVMDIKLQLKQSSVQVNAAGPLAVALADTNTGRYPDNVYKLLKEYFRQFVAC 1900
Qy 2028 GQALVNEBELIKEDOLEYOEBKANKYREMAKELSEIMHQICPLEKERTSVLPNSLIHFA 2087
Db 1901 GQALVNEBELIKEDOLEYOEBKANKYREMAKELSEIMHQICPLEKERTSVLPNSLIHFA 1960
```

```
Qy 2088 ISGTFSTVMHGMTSSSSSV 2107
Db 1961 ISGTFSTVMHGMTSSSSSV 1980

RESULT 4
US-09-736-969A-94
; Sequence 94, Application US/09736969A
; Patent No. US20020068302A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-4 Transmembrane Protein
; FILE REFERENCE: 020054-000411US
; CURRENT APPLICATION NUMBER: US/09/736,969A
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 1980
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-2
US-09-736-969A-94

Query Match 90.8%; Score 9926.5; DB 10; Length 1980;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;

Qy 73 MLFPYDPOFRAIILAROGRYICSTVPARAEBAOSLFTECKITYSMDHLVNYREDS 132
Db 1 MLFPYDPOFRAIILAROGRYICSTVPARAEBAOSLFTECKITYSMDHLVNYREDS 60
Qy 133 GEFROLPNKVVLDLCPVHYVEVDSEVDKEDPAASLGSGKGITGIGWLKGNMNSAISV 192
Db 61 GEFROLPNKVVLDLCPVHYVEVDSEVDKEDPAASLGSGKGITGIGWLKGNMNSAISV 120
Qy 193 TWRSPKRRPFHLIQGDSGYKFEFLKD--LQKEPKGIFLGFVGVSPRNKVRFPAPFL 250
Db 121 TWRSPKRRPFHLIQGDSGYKFEFLKD--LQKEPKGIFLGFVGVSPRNKVRFPAPFL 179
```



251 KNOOKSSYLLAADSEVEMERITILNKILQNEEAMOEKRNDSHEDDEOSKLEGGSG 310  
 180 KNOOKSSYLLAADSEVEMERITILNKILQNEEAMOEKRNDSHEDDEOSKLEGGSG 239  
 311 LQSYBELAKSAREABIKLKSESRYKLYLDPDAQKLDPSASAPVKSFEKKGKILVK 370  
 240 LQSYBELAKSAREABIKLKSESRYKLYLDPDAQKLDPSASAPVKSFEKKGKILVK 299  
 371 CNDLSNLOCCVAENEEGPTTNEPFFVTLSPDICTNKRKISADPHVDLNHPSVRQMT 430  
 300 CNDLSNLOCCVAENEEGPTTNEPFFVTLSPDICTNKRKISADPHVDLNHPSVRQMT 359  
 431 TSPALNNGSGPETOSALRGILHEAAMQYKOGIFSTVCPHPIFLVARIKYLQGSITHC 490  
 360 TSPALNNGSG-GSPSYLKGILHEAAMQYKOGIFSTVCPHPIFLVARIKYLQGSITHC 418  
 491 AEPYKSSDSKVAQKVLKNAKQACORLQYRMPFMAARTLEPKASGNLDKNAKPSATY 550  
 419 AEPYKSSDSKVAQKVLKNAKQACORLQYRMPFMAARTLEPKASGNLDKNAKPSATY 478  
 551 RQDSNKLSDMDMLKILADPRKPEKAKLPIYLGNLDTIDNVSDEPNVNSSYITTKOP 610  
 479 RQDSNKLSDMDMLKILADPRKPEKAKLPIYLGNLDTIDNVSDEPNVNSSYITTKOP 538  
 611 ETCSKPTITEVEBEFVPCIPKHTQPYTIYNHLVYVPKYLKQOSKSPAKARIALCIEF 670  
 539 ETCSKPTITEVEBEFVPCIPKHTQPYTIYNHLVYVPKYLKQOSKSPAKARIALCIEF 598  
 671 KQDEEDSOPKLCTYRPGGPFVTRSAFAVLHHQNEPEFYDEIKIELPTQLEKHHLL 730  
 599 KQDEEDSOPKLCTYRPGGPFVTRSAFAVLHHQNEPEFYDEIKIELPTQLEKHHLL 658  
 731 TEFHVSQDSSKSTKKRDVETQVGYSWLPLKDRVVTSEQHIVSANTLPGHLYG 790  
 659 TEFHVSQDSSKSTKKRDVETQVGYSWLPLKDRVVTSEQHIVSANTLPGHLYG 718  
 791 LGMGRHYGPEIKVNDGKPLKISTHVSIVYQDQHLNFPQYCKTSSGAQALNELY 850  
 719 LGMGRHYGPEIKVNDGKPLKISTHVSIVYQDQHLNFPQYCKTSSGAQALNELY 778  
 851 KYLKSIAHEGHWMIAPLPTILNQLFRVLTARATQEEVAANVTREVIHHVAAQCEBGLSH 910  
 779 KYLKSIAHEGHWMIAPLPTILNQLFRVLTARATQEEVAANVTREVIHHVAAQCEBGLSH 838  
 911 LBSYVYKAAKABEYVASEKTYHEELTKGMTILKRSADPLSNKLKXSWFFDVLIS 970  
 839 LBSYVYKAAKABEYVASEKTYHEELTKGMTILKRSADPLSNKLKXSWFFDVLIS 898  
 971 MAQHLIENSKVLLRNQRPASVYHAAVEVYVNMAMPITQKPRDNPEASKNANHSIAVFI 1030  
 899 MAQHLIENSKVLLRNQRPASVYHAAVEVYVNMAMPITQKPRDNPEASKNANHSIAVFI 958  
 1031 KQCFITMDRGVFKQJNNYISCFAPGDPKTLFEYKFEPLRVVNCHEHYIPLNLPMPFGK 1090  
 959 KQCFITMDRGVFKQJNNYISCFAPGDPKTLFEYKFEPLRVVNCHEHYIPLNLPMPFGK 1018  
 1091 RIORVODLQDLSLTDEFCRNHFLVGLLREVTALQERREVLIAISVLKXLLHGSFD 1150  
 1019 RIORVODLQDLSLTDEFCRNHFLVGLLREVTALQERREVLIAISVLKXLLHGSFD 1078  
 1151 DRYASRQARIALTLPLFGLLIENVQRIINVDVPPFNAGMTVKDESIALPVPNPV 1210  
 1079 DRYASRQARIALTLPLFGLLIENVQRIINVDVPPFNAGMTVKDESIALPVPNPV 1138  
 1211 TPQKSTLQDLSLHKLGAISGIASPYTTSTPNINSVRNADSGSLISTDSGNSLPERNS 1270  
 1139 TPQKSTLQDLSLHKLGAISGIASPYTTSTPNINSVRNADSGSLISTDSGNSLPERNS 1198  
 1271 EKSNSLDKXQSSSTLGNVVRCDKLDQSEIKSLMLCFLYILKMSMDALFTYWNKASTSE 1330  
 1199 EKSNSLDKXQSSSTLGNVVRCDKLDQSEIKSLMLCFLYILKMSMDALFTYWNKASTSE 1258

1331 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1367  
 1259 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1318  
 1368 SIDNSLTENHSGYSHDADVLHOSLLEANIATEVCLTALDTSLSFTLAFKNOILLADHGNP 1427  
 1319 SIDNSLTENHSGYSHDADVLHOSLLEANIATEVCLTALDTSLSFTLAFKNOILLADHGNP 1378  
 1428 LMKXVFDVYLCFLOHGOSETLXONVFTLRSLIYFPESTFEGRADMCALCYELTKCCN 1487  
 1379 LMKXVFDVYLCFLOHGOSETLXONVFTLRSLIYFPESTFEGRADMCALCYELTKCCN 1438  
 1488 SKLSSIRTEASQLYFLMRNNPDYTGKSFVTHLQVLISSQLIADVYLGIGTFPOSL 1547  
 1439 SKLSSIRTEASQLYFLMRNNPDYTGKSFVTHLQVLISSQLIADVYLGIGTFPOSL 1498  
 1499 SIINNCAUSDRLIKTSSFSDDVKTCKRIFTVLAATAQKEHENDPEMLVDLQYSLASY 1558  
 1608 ASTPELRKTLQDSMARIVHKNQDLSEAMCYVHTALVAEYLTRKNAVQWEPPLPHSHS 1667  
 1559 ASTPELRKTLQDSMARIVHKNQDLSEAMCYVHTALVAEYLTRKNAVQWEPPLPHSHS 1603  
 1668 ACLRRSGGVFROGCTAFRVTIPNIDEASMMEDVGMQVHNEEDVLMELLEQCADGLMK 1727  
 1604 -----GVFRQGTAFRVTIPNIDEASMMEDVGMQVHNEEDVLMELLEQCADGLMK 1655  
 1728 AERYELADIYKLLIPIYERKRDPERLAHLDTLHRAVSKTEVHSGRRLIGTYFRVAF 1787  
 1656 AERYELADIYKLLIPIYERKRDPERLAHLDTLHRAVSKTEVHSGRRLIGTYFRVAF 1678  
 1788 FQGAAYQVFTSETEVEGFEDEDEKEXITYEKPXLTPLSEISORLLKLYSDKFGSENVYK 1847  
 1679 -----FPEDEDEKEXITYEKPXLTPLSEISORLLKLYSDKFGSENVYK 1720  
 1848 IODSGKVPKQDLSKXAYIOVTHVIPFDEKELQERKTEPERSHNIIRFMEPMTOTGK 1907  
 1721 IODSGKVPKQDLSKXAYIOVTHVIPFDEKELQERKTEPERSHNIIRFMEPMTOTGK 1780  
 1908 ROGGYBEOCKRTIITAIHCPYVKKRIPWYQHTDNLNPEVAIDEMSKYAEIRPOLCS 1967  
 1781 ROGGYBEOCKRTIITAIHCPYVKKRIPWYQHTDNLNPEVAIDEMSKYAEIRPOLCS 1840  
 1968 SAEVDMIKLQKLGQSVSYVNAAGPLAYARAFLDNTKRYVDNRYKLLKEVFRQFVEAC 2027  
 1841 SAEVDMIKLQKLGQSVSYVNAAGPLAYARAFLDNTKRYVDNRYKLLKEVFRQFVEAC 1900  
 2028 GQALAVNERLIKEDOLEYQEBKANYREMAKELSEIMHEOICPLEBKTSLVPSLHIFNA 2087  
 1901 GQALAVNERLIKEDOLEYQEBKANYREMAKELSEIMHEOICPLEBKTSLVPSLHIFNA 1960  
 2088 ISGTPSTVHGMSTSSSVY 2107  
 1961 ISGTPSTVHGMSTSSSVY 1980

RESULT 5  
 US-09-736-960-91  
 ; Sequence 91, Application US/09736960  
 ; Patent No. US20020102267A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lu, Peter  
 ; APPLICANT: Garman, Jonathan David  
 ; APPLICANT: Candia III, Albert Frederick  
 ; APPLICANT: Athor Vita Corporation  
 ; TITLE OF INVENTION: CLASP-5 Transmembrane Protein  
 ; FILE REFERENCE: 020054-000511US  
 ; CURRENT APPLICATION NUMBER: US/09/736,960  
 ; CURRENT FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: US 60/160,860  
 ; PRIOR FILING DATE: 1999-10-21  
 ; PRIOR APPLICATION NUMBER: US 60/162,498

/ PRIOR FILING DATE: 1999-10-29  
/ PRIOR APPLICATION NUMBER: US 60/170,453  
/ PRIOR FILING DATE: 1999-12-13  
/ PRIOR APPLICATION NUMBER: US 60/176,195  
/ PRIOR FILING DATE: 2000-01-14  
/ PRIOR APPLICATION NUMBER: US 60/182,296  
/ PRIOR FILING DATE: 2000-02-14  
/ PRIOR APPLICATION NUMBER: US 09/547,276  
/ PRIOR FILING DATE: 2000-04-11  
/ PRIOR APPLICATION NUMBER: US 60/196,267  
/ PRIOR FILING DATE: 2000-04-11  
/ PRIOR APPLICATION NUMBER: US 60/196,460  
/ PRIOR FILING DATE: 2000-04-11  
/ PRIOR APPLICATION NUMBER: US 60/196,527  
/ PRIOR FILING DATE: 2000-04-11  
/ PRIOR APPLICATION NUMBER: US 60/196,528  
/ PRIOR FILING DATE: 2000-04-11  
/ PRIOR APPLICATION NUMBER: US 09/687,837  
/ PRIOR FILING DATE: 2000-10-13  
/ PRIOR APPLICATION NUMBER: US 60/240,503  
/ PRIOR FILING DATE: 2000-10-13  
/ PRIOR APPLICATION NUMBER: US 60/240,508  
/ PRIOR FILING DATE: 2000-10-13  
/ PRIOR APPLICATION NUMBER: US 60/240,539  
/ PRIOR FILING DATE: 2000-10-13  
/ PRIOR APPLICATION NUMBER: US 60/240,543  
/ PRIOR FILING DATE: 2000-10-13  
/ NUMBER OF SEQ ID NOS: 134  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 91  
/ LENGTH: 1980  
/ TYPE: PR1  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: human CLASP-2  
US-09-736-960-91

Query Match 90.8%; Score 9926.5; DB 10; Length 1980;  
Best Local Similarity 93.8%; Pred. No. 0;  
Matches 1992; Conservative 8; Mismatches 15; Indels 105; Gaps 6;  
73 MLFPYDFQIALRQGRYICSTVPAAKEEAQSLFVTECITKYNDSMHLVNYKEDYS 132  
1 MLFPYDFQIALRQGRYICSTVPAAKEEAQSLFVTECITKYNDSMHLVNYKEDYS 60  
133 GEFROLPNVYVLDKLPVYVYEVDEVEDDEDAASIGSGKGTTRKGMLYKGNMSAISV 192  
61 GEFROLPNVYVLDKLPVYVYEVDEVEDDEDAASIGSGKGTTRKGMLYKGNMSAISV 120  
193 TRSRFRRFFHLIQLDGSYKPEFLKD--LOKEPKGSIFLGFLYGVSPNNKVRRAFAEL 250  
121 TRSRFRRFFHLIQLDGSYVNLNFYDEKISKEPKSIFLDSGMGV-VQNNKYRRRAFAEL 179  
251 KNQDSSYLLADSEVEMEWITLTKLILQLNFEAAMQIKRNGDSHEDDEQSLDESGSG 310  
180 KNQDSSYLLADSEVEMEWITLTKLILQLNFEAAMQIKRNGDSHEDDEQSLDESGSG 239  
311 LDSYLPKAKSAREATIKLKSSGRVLYFLDPDAQCLDSSAEPYKSTREKRGKILYK 370  
240 LDSYLPKAKSAREATIKLKSSGRVLYFLDPDAQCLDSSAEPYKSTREKRGKILYK 299  
371 CNDLSFNLOCCVAENEGPTTNVEPFVTLSLFDIKYKRIKISADPFVLDLNFHSVRQMIAT 430  
300 CNDLSFNLOCCVAENEGPTTNVEPFVTLSLFDIKYKRIKISADPFVLDLNFHSVRQMIAT 359  
431 TSPALMNGSGPETQSALRGILHEAAMQYKQGISVTCCHPDIFLVAIEKYLQGSITHC 490  
360 TSPALMNGSG--QSPSYLKGILHEAAMQYKQGISVTCCHPDIFLVAIEKYLQGSITHC 418  
491 ASPYKSSDSKVAQVLAQNAKQACORLQYRMPFAMARATLTKDASGULDKNARPSATY 550  
419 ASPYKSSDSKVAQVLAQNAKQACORLQYRMPFAMARATLTKDASGULDKNARPSATY 478

QY 551 RODSNKLSDNDMLKLLADFRKPKAKLPEVILGNLDTITDYNVSDPFNNVSSSYPTKQF 610  
DB 479 RODSNKLSDNDMLKLLADFRKPKAKLPEVILGNLDTITDYNVSDPFNNVSSSYPTKQF 538  
QY 611 ETCSKTPTTTEVEEVPICPKATQYTYITNHLVYYPKYLKDSQSKPAKANIALCIF 670  
DB 539 ETCSKTPTTTEVEEVPICPKATQYTYITNHLVYYPKYLKDSQSKPAKANIALCIF 598  
QY 671 KDSDEDSQPLKCIIGRPGCPVFTSAFAVAHHQNPFEYDIEKLEPTOLHEKHLL 730  
DB 599 KDSDEDSQPLKCIIGRPGCPVFTSAFAVAHHQNPFEYDIEKLEPTOLHEKHLL 658  
QY 731 TFFHVSQDSSKSGSTRKDDVETQVGSWLPKDKGRVTSQOH1PVSANLPSGYLGYOE 790  
DB 659 TFFHVSQDSSKSGSTRKDDVETQVGSWLPKDKGRVTSQOH1PVSANLPSGYLGYOE 718  
QY 791 LQNGHRYGPEIKWVDGKPLKISTHLYSVTVDQOHLNPFQYCKTESGAQALGNEV 850  
DB 739 LQNGHRYGPEIKWVDGKPLKISTHLYSVTVDQOHLNPFQYCKTESGAQALGNEV 778  
QY 851 KYLKSILHMEGHVMTAFLPTILNOLFRVLTTRATQSEVAVNVRVYIHHVAQHEBLSH 910  
DB 779 KYLKSILHMEGHVMTAFLPTILNOLFRVLTTRATQSEVAVNVRVYIHHVAQHEBLSH 838  
QY 911 LRSYKAYKAPYVYASEKYVHEELTKSMITILKPSADPLTSNKLKYSWFFPDVILKS 970  
DB 839 LRSYKAYKAPYVYASEKYVHEELTKSMITILKPSADPLTSNKLKYSWFFPDVILKS 898  
QY 971 MAQHLIENSKVLLNORFPASVYHHAETVNMMLPHITQKRDNPBASKANHSLAVY 1030  
DB 899 MAQHLIENSKVLLNORFPASVYHHAETVNMMLPHITQKRDNPBASKANHSLAVY 958  
QY 1031 KRCSFTFMRGPFVKQIINNYISCFAPGDPKTLPEYKPEFLRVYCNHEHY1PLNLPMPFGG 1090  
DB 959 KRCSFTFMRGPFVKQIINNYISCFAPGDPKTLPEYKPEFLRVYCNHEHY1PLNLPMPFGG 1018  
QY 1091 RIQRYODQLDYSLTDEFCRNHFLVGLLREVGTALQEFREVRLLAISVLKULLIKHSPD 1150  
DB 1019 RIQRYODQLDYSLTDEFCRNHFLVGLLREVGTALQEFREVRLLAISVLKULLIKHSPD 1078  
QY 1151 DRYASRSHQARIATYLPFLGILLIENVQRIINRDVSPFPVNAQMTYKBSLALPAVNPV 1210  
DB 1079 DRYASRSHQARIATYLPFLGILLIENVQRIINRDVSPFPVNAQMTYKBSLALPAVNPV 1138  
QY 1211 TPQKSTLDSNLHKOLLGASIGIASPYTSTPNINSVRNADRGSLISTDSGNSLPERNS 1270  
DB 1139 TPQKSTLDSNLHKOLLGASIGIASPYTSTPNINSVRNADRGSLISTDSGNSLPERNS 1198  
QY 1271 EKSNSLDKHOQSSSTLGNISVYRCDKLDQSEIKSLMCFYIILKSMGSDALFTYMNKASTSE 1330  
DB 1199 EKSNSLDKHOQSSSTLGNISVYRCDKLDQSEIKSLMCFYIILKSMGSDALFTYMNKASTSE 1258  
QY 1331 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1367  
DB 1259 LMDPFTISEVCLHOFQYMGKRYIAHQEGLGIVHDKRSQTLPVSHNRGMMHARLQOLG 1318  
QY 1368 SLDNLSTPHNSYGHSDADAVLHQSLSLEANTATVCLTALDITSLFTLAFNQULLADHGNP 1427  
DB 1339 SLDNLSTPHNSYGHSDADAVLHQSLSLEANTATVCLTALDITSLFTLAFNQULLADHGNP 1378  
QY 1428 LMKKVPDYVLCFLQHQSEETALKVFTALRSILYFSPSTFYEGRADMCALCYEILKCN 1487  
DB 1379 LMKKVPDYVLCFLQHQSEETALKVFTALRSILYFSPSTFYEGRADMCALCYEILKCN 1438  
QY 1488 SKLSSIRTEASQLLFLKRNNDYTGKGSFVATHLQVILISQULLADVVGISGTRFOQSL 1547  
DB 1439 SKLSSIRTEASQLLFLKRNNDYTGKGSFVATHLQVILISQULLADVVGISGTRFOQSL 1498  
QY 1548 SIINNCANDRLIKHTSSSDVKDLTKRIRATYMTAOMKEHENPBMVLVDIQLYSAXSY 1607  
DB 1499 SIINNCANDRLIKHTSSSDVKDLTKRIRATYMTAOMKEHENPBMVLVDIQLYSAXSY 1558  
QY 1608 ASTPELRKTWLDMSARIHVKNQDLSBAAMCYVHTALVAEYLTTRKEAVQWBPPLPHSHS 1667

```
|||||
Db 1559 ASTPELRKTKWLDMSMAIHVKNGLDSEAAKCYHVHTALVAEYLTRK----- 1603
Qy 1668 ACIARRRGVFRGCGTAFRVITPNIDEBSMMEVCMODVPHMEDLMELLECAGDGLK 1727
Db 1604 -----GVFRGCGTAFRVITPNIDEBSMMEVCMODVPHMEDLMELLECAGDGLK 1655
Qy 1728 AERYELIADYKLIIEYERKDPFERLADLTILHRAYSKVTEVMSGRLLGTFRVAF 1787
Db 1656 AERYELIADYKLIIEYERKDPFERLADLTILHRAYSKVTEVMSGRLLGTFRVAF 1678
Qy 1788 FGGAAOYQFTDSTDEVEGFEDEDGKEYIYKEPKLTPLSISQRLIKLYSDKFGSNVAK 1847
Db 1679 -----FPEDEDGKEYIYKEPKLTPLSISQRLIKLYSDKFGSNVAK 1720
Qy 1848 IODSGKVPKDLSDSKAYIQVTHVIPPEPEKEIOERKTEFEERSHNIIRFEMEPFTQTK 1907
Db 1721 IODSGKVPKDLSDSKAYIQVTHVIPPEPEKEIOERKTEFEERSHNIIRFEMEPFTQTK 1780
Qy 1908 ROGVEEQCKRRTILTAIHCFPVYKKRIPMYQHHTDLNPIEVAIDEMSKVAELRQLCS 1967
Db 1781 ROGVEEQCKRRTILTAIHCFPVYKKRIPMYQHHTDLNPIEVAIDEMSKVAELRQLCS 1840
Qy 1968 SAEVDMIKQLKQSSVSQVNAAPLAYARAFLLDNTNRYKYPDNKVKLLKEVFRQVEAC 2027
Db 1841 SAEVDMIKQLKQSSVSQVNAAPLAYARAFLLDNTNRYKYPDNKVKLLKEVFRQVEAC 1900
Qy 2028 GQALANERLIKEDOLEYOEMKANYREMAKELSEIMHEQICPLEKTSYLPNSLIHFA 2087
Db 1901 GQALANERLIKEDOLEYOEMKANYREMAKELSEIMHEQICPLEKTSYLPNSLIHFA 1960
Qy 2088 ISGTPSTWVHGHTSSSSSVY 2107
Db 1961 ISGTPSTWVHGHTSSSSSVY 1980

RESULT 6
US-09-978-244A-27
; Sequence 27, Application US/0978244A
; Publication No. US20030103992A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter S
; APPLICANT: Garman, Jonathan D.
; APPLICANT: Candia III, Albert F.
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
; FILE REFERENCE: 020554-000161US
; CURRENT APPLICATION NUMBER: US/09/978, 244A
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/310,028
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/737,246
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,969
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,960
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,968
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/240,545
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
```

```
; LENGTH: 2040
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CLASP-2
US-09-978-244A-27

Query Match 90.4%; Score 9886.5; DB 9; Length 2040;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 165; Gaps 7;

Qy 73 MLFPYDDFOIALRRQGRYICSTVPAAKEBAQSLFV----- 110
Db 1 MLFPYDDFOIALRRQGRYICSTVPAAKEBAQSLFVTECIKTYSNDWHLVYKYEDYS 60
Qy 111 -----TECIKTYSNDWHLVYKYEDYS 132
Db 61 GEFROLPNKVPKLDKLPVHVYEVDEADKEDDAASLGSTECIKTYNSDWHLVNYKYEDYS 120
Qy 133 GEFROLPNKVPKLDKLPVHVYEVDEADKEDDAASLGSGITGHGWLKGNMNSAISV 192
Db 121 GEFROLPNKVPKLDKLPVHVYEVDEADKEDDAASLGSGITGHGWLKGNMNSAISV 180
Qy 193 TWRSEKRRFFHLIQLDGGSYKFEFLKD--LQKEPKGSIFLGFLYGVSPRNNKVRRAPEFL 250
Db 181 TWRSEKRRFFHLIQLDGGSYKFEFLKD--LQKEPKGSIFLGFLYGVSPRNNKVRRAPEFL 239
Qy 251 KNODSSYLLADSEVMEEMWITLNLKILQNFEAAMQEKRRGDSHEDDEOSKLGSGSG 310
Db 240 KNODSSYLLADSEVMEEMWITLNLKILQNFEAAMQEKRRGDSHEDDEOSKLGSGSG 299
Qy 311 LDSYLPBELAKSAREAEIKLKSBSRYKLFLYDPAQKLPFSSAEPYKSFEEKFGKRIIVK 370
Db 300 LDSYLPBELAKSAREAEIKLKSBSRYKLFLYDPAQKLPFSSAEPYKSFEEKFGKRIIVK 359
Qy 371 CNDLSFNLQCCVAENEBEGFTNVBEPFVTLSPFDIKYNNKISADPHVDLHNSVROMIAT 430
Db 360 CNDLSFNLQCCVAENEBEGFTNVBEPFVTLSPFDIKYNNKISADPHVDLHNSVROMIAT 419
Qy 431 TSPALMNGSGPBTOSALGILHEAAMOYPKQISFVTCGHPPIFLVAREKYLQGSITHC 490
Db 420 TSPALMNGSG-OSPEVLKGLHEAAMOYPKQISFVTCGHPPIFLVAREKYLQGSITHC 478
Qy 491 AEPYKSSDSSKVAOKVLAQAKACORLQGYEMPRPAMARTLFXDASGLDNKARFSAY 550
Db 479 AEPYKSSDSSKVAOKVLAQAKACORLQGYEMPRPAMARTLFXDASGLDNKARFSAY 538
Qy 551 RQDSNKLSDMDLKLADFRKPEKRAKLPVILGNLIDITIDNVSDFPNVNSSYIPTKQF 610
Db 539 RQDSNKLSDMDLKLADFRKPEKRAKLPVILGNLIDITIDNVSDFPNVNSSYIPTKQF 598
Qy 611 ETCSKTPITFEVEEVEVPCLPKHTOPYTYITNHLVYYPKLYKDSQSPKAKANIMACIEF 670
Db 599 ETCSKTPITFEVEEVEVPCLPKHTOPYTYITNHLVYYPKLYKDSQSPKAKANIMACIEF 658
Qy 671 KQDSEDSQPLKCIYGRPGPVFTSAPAAVLHHHQNBEFYDEIKIELPTQJLHEKHLL 730
Db 659 KQDSEDSQPLKCIYGRPGPVFTSAPAAVLHHHQNBEFYDEIKIELPTQJLHEKHLL 718
Qy 731 TFFHVSQDSSKSGSTKKDDVETQVGYSWLPLKQGRVYTSQHI PVSANLPSGYLYGYOE 790
Db 719 TFFHVSQDSSKSGSTKKDDVETQVGYSWLPLKQGRVYTSQHI PVSANLPSGYLYGYOE 778
Qy 791 LQMGHRYGPEIKWNGGKPLKISTHLVSTVYTOQHLLHNFQYQOKTESGQALGNELV 850
Db 779 LQMGHRYGPEIKWNGGKPLKISTHLVSTVYTOQHLLHNFQYQOKTESGQALGNELV 838
Qy 851 KTLKSLHMEGHVMIAPLPTILNQFLVLTTRATOEEVAVNVTYVLIHHVAAQCHEGLESJ 910
Db 839 KTLKSLHMEGHVMIAPLPTILNQFLVLTTRATOEEVAVNVTYVLIHHVAAQCHEGLESJ 898
Qy 911 LRSYKAYAKAPYVASSEYKTVHEELTKSMTTILKPSADFLTSNKLKYSNFFDVLIKS 970
Db 899 LRSYKAYAKAPYVASSEYKTVHEELTKSMTTILKPSADFLTSNKLKYSNFFDVLIKS 878
```

```

Db 899 LRSYKAYAKBPPYVASEYKTVHBEJLTKSMITLTKRSADPLTSLNKLRLRSWFFDVLKLS 958
Qy 971 MAOHLIENSKVTLRLNORPPASVYHAVETVYVNMMLPHITQKFRDNEPASKNHLAVPI 1030
Db 959 MAOHLIENSKVTLRLNORPPASVYHAAETVYVNMMLPHITQKGDNEPASKNHLAVPI 1018
Qy 1031 KSCFTFMDRGPFVKQJNNYISCPAPDPTLFEYKEFLRVYCNHEHY1PLMLPMPFGKG 1090
Db 1019 KSCFTFMDRGPFVKQJNNYISCPAPDPTLFEYKEFLRVYCNHEHY1PLMLPMPFGKG 1078
Qy 1091 RIQRVODLDVSLDEFCRNHPVLGGLREVGTALQEFREVRLLAISVLKULLIKHSD 1150
Db 1079 RIQRVODLDVSLDEFCRNHPVLGGLREVGTALQEFREVRLLAISVLKULLIKHSD 1138
Qy 1151 DRYASRSHOARIATYLPFGLLIENVORINRDVSPFVNMGMVYKDESLAPVNPVY 1210
Db 1139 DRYASRSHOARIATYLPFGLLIENVORINRDVSPFVNMGMVYKDESLAPVNPVY 1198
Qy 1211 TPQKSTLDNSLHKDLGAIISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNS 1270
Db 1199 TPQKSTLDNSLHKDLGAIISGIASPYTTSTPNINSVRNADSRGSLISTDSGNSLPERNS 1258
Qy 1271 EKSNSLDRKQOSTLGNSVVRCDKLDQSEIKSLMCFYILKMSDDALFTYWNKASTSE 1330
Db 1259 EKSNSLDRKQOSTLGNSVVRCDKLDQSEIKSLMCFYILKMSDDALFTYWNKASTSE 1318
Qy 1331 LMDPFTISFVCLHOPQYMGKRYIA-----RTGMHARLQOLG 1367
Db 1319 LMDPFTISFVCLHOPQYMGKRYIARNOBGLGPIVHDKRSQTLPVSNRNGMMHARLQOLG 1378
Qy 1368 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDITLSLFTLAFRNOULLADHGNP 1427
Db 1379 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDITLSLFTLAFRNOULLADHGNP 1438
Qy 1428 LMKKVPDYVLCFLQKQOSTALKNVFTALRSLLYKPESTFYEGRADMCALCYEILKCN 1487
Db 1439 LMKKVPDYVLCFLQKQOSTALKNVFTALRSLLYKPESTFYEGRADMCALCYEILKCN 1498
Qy 1488 SKLSIRTEASQLVLYLMKNPNFYTGKSFVPTHLOVITISVSQLADVVIGIGSTRFOQL 1547
Db 1499 SKLSIRTEASQLVLYLMKNPNFYTGKSFVPTHLOVITISVSQLADVVIGIGSTRFOQL 1558
Qy 1548 SIINNCAANDRLIKHTSFSSDVLDLTKIRITVLMATAQKHEHNDPEMLVDLYQSLAKSY 1607
Db 1559 SIINNCAANDRLIKHTSFSSDVLDLTKIRITVLMATAQKHEHNDPEMLVDLYQSLAKSY 1618
Qy 1608 ASTPELRKTWLDSMARIHYNGDLSBANACYHTALVAEYLTRKEAVOMBEPPLPHSHS 1667
Db 1639 ASTPELRKTWLDSMARIHYNGDLSBANACYHTALVAEYLTRK----- 1663
Qy 1668 ACLRSRSGVPRQCGTAPFVITPNIDEASMMEDVGMQVHNEDVLMELBQCADGLMK 1727
Db 1664 -----GVPRQCGTAPFVITPNIDEASMMEDVGMQVHNEDVLMELBQCADGLMK 1715
Qy 1728 AERYELIADIYKLIIPYIEKRRDPERLAHLVDTLHRAVEKTVEMHSGRLLIGTYRVAV 1787
Db 1716 AERYELIADIYKLIIPYIEKRRD----- 1738
Qy 1788 FQQAAYOYFTDSETVDEGFPEDDEGKEYIYKEPKLTPLAEISORLIKLYSDKFGSENVM 1847
Db 1739 -----FPEDEDEGKEYIYKEPKLTPLAEISORLIKLYSDKFGSENVM 1780
Qy 1848 IODSGKVNPKDLDSKAYAYIOVHVIPFPPEKELQEKTKSPESHNRIRRMFEMPFTQTK 1907
Db 1781 IODSGKVNPKDLDSKAYAYIOVHVIPFPPEKELQEKTKSPESHNRIRRMFEMPFTQTK 1840
Qy 1908 ROGQVEEQCKRRITLTAHCPCPYVKRIPVMTQHTDNLPIEVALIDEMSKYAEALRQLCS 1967
Db 1841 ROGQVEEQCKRRITLTAHCPCPYVKRIPVMTQHTDNLPIEVALIDEMSKYAEALRQLCS 1900
Qy 1968 SAEVDMIKYQLKQGSVSVQVNAAGPLAYARAFLDDNTNTRYIPDNKYLKXVFRQVEAC 2027
Db 1901 SAEVDMIKYQLKQGSVSVQVNAAGPLAYARAFLDDNTNTRYIPDNKYLKXVFRQVEAC 1960

```

```

Qy 2028 GOALAVNERLLIKEDQLEYQEBKANYREMAKELSEIMHEQICPLEEKTSVLPNSLIHENA 2087
Db 1961 GOALAVNERLLIKEDQLEYQEBKANYREMAKELSEIMHEQICPLEEKTSVLPNSLIHENA 2020
Qy 2088 ISGTPTSTVHGHTSSSSSV 2107
Db 2021 ISGTPTSTVHGHTSSSSSV 2040

RESULT 7
US-09-736-968A-10
/ Sequence 10, Application US/09736968A
/ Patent No. US20020169283A1
/ GENERAL INFORMATION:
/ APPLICANT: Lu, Peter
/ APPLICANT: Garman, Jonathan David
/ APPLICANT: Candia III, Albert Frederick
/ APPLICANT: Arbor Vita Corporation
/ TITLE OF INVENTION: CLASP-7 Transmembrane Protein
/ FILE REFERENCE: 020054-000611US
/ CURRENT APPLICATION NUMBER: US/09/736,968A
/ CURRENT FILING DATE: 2000-12-13
/ PRIOR APPLICATION NUMBER: US 60/160,860
/ PRIOR FILING DATE: 1999-10-21
/ PRIOR APPLICATION NUMBER: US 60/162,498
/ PRIOR FILING DATE: 1999-10-29
/ PRIOR APPLICATION NUMBER: US 60/170,453
/ PRIOR FILING DATE: 1999-12-13
/ PRIOR APPLICATION NUMBER: US 60/176,195
/ PRIOR FILING DATE: 2000-01-14
/ PRIOR APPLICATION NUMBER: US 60/182,296
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: US 09/547,276
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,267
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,460
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,527
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,528
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 09/687,837
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,503
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,508
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,539
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,543
/ PRIOR FILING DATE: 2000-10-13
/ SOFTWARE: Patent In Ver. 2.1
/ NUMBER OF SEQ ID NOS: 115
/ SEQ ID NO 10
/ LENGTH: 1534
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: human KIAA1058
US-09-736-968A-10

Query Match 72.9%; Score 7976; DB 9; Length 1534;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 536 ASGNLDKARARSATYRQDSNKLSDNDMLKTLADFPKPEKMAKLPVILGNLDTITDINVSDD 595
Db 1 ASGNLDKARARSATYRQDSNKLSDNDMLKTLADFPKPEKMAKLPVILGNLDTITDINVSDD 60
Qy 596 PPNVNSSYIPTKQETCKPTTFEEVEFPVCPIDKATQPYTTIYNHLVYVPKYKYSQ 655

```

```

Db      61  PBYVNSYIPTKQFETCSKTPITFEVEBEVPCIPKHTOPYTITNHLVYVPCILKYDSQ  120
Qy      656  KSPAKARNIACIEFDOSDEDSQPLKCIYGRPGGVPTRSAPAAVLHHHONBEYDEIK  715
Db      121  KSPAKARNIACIEFDOSDEDSQPLKCIYGRPGGVPTRSAPAAVLHHHONBEYDEIK  180
Qy      716  IBLPTQHEKHLHLTFPHVSCDSSKSGTKRDVETOVGVSWPLKXGRVVTSEQHI  775
Db      181  IBLPTQHEKHLHLTFPHVSCDSSKSGTKRDVETOVGVSWPLKXGRVVTSEQHI  240
Qy      776  PVSANIPSGYLQVQELGMRHYPEIKMDGKPLKISTHLVSTYTYTODOLHNFQYC  835
Db      241  PVSANIPSGYLQVQELGMRHYPEIKMDGKPLKISTHLVSTYTYTODOLHNFQYC  300
Qy      836  OKTESGAOLAGNELVYKLSLHMEGHVMIAPLTLINOLFVLTATOEVAVANTRTY  895
Db      301  OKTESGAOLAGNELVYKLSLHMEGHVMIAPLTLINOLFVLTATOEVAVANTRTY  360
Qy      896  IHVVAOCHEGLESRLSYVKAAPYVASEYKTVHEBLTKSMITLIKPSADFLTSNK  955
Db      361  IHVVAOCHEGLESRLSYVKAAPYVASEYKTVHEBLTKSMITLIKPSADFLTSNK  420
Qy      956  LKYSWFFFDVLIKSAQHLIENSKYKLRNORFPASSTHVAETVYVNMMPHTQKFRN  1015
Db      421  LKYSWFFFDVLIKSAQHLIENSKYKLRNORFPASSTHVAETVYVNMMPHTQKFRN  480
Qy      1016  PEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLFEYKFEFLAVCNH  1075
Db      481  PEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLFEYKFEFLAVCNH  540
Qy      1076  EHYIPLNLEMPFGKRIQRYQDLQDYSLTDEFCRNHFLVGLLREBVGALQEFREVLI  1135
Db      541  EHYIPLNLEMPFGKRIQRYQDLQDYSLTDEFCRNHFLVGLLREBVGALQEFREVLI  600
Qy      1136  AISVLKNLIIKHSFDRVYASRSHQAIATLYLPFELLLENQRIWVRVSPPVNAAGT  1195
Db      601  AISVLKNLIIKHSFDRVYASRSHQAIATLYLPFELLLENQRIWVRVSPPVNAAGT  660
Qy      1196  VDESLALPAVNPVLPQKSTLDNSLHKDLGASIPYTTSTPNINSVNRNADSRGS  1255
Db      661  VDESLALPAVNPVLPQKSTLDNSLHKDLGASIPYTTSTPNINSVNRNADSRGS  720
Qy      1256  LISTDSGNSLPERNSEKNSLDKQOQSTLQNSVVRCDKLDOSEIKSLMCFLYILKNS  1315
Db      721  LISTDSGNSLPERNSEKNSLDKQOQSTLQNSVVRCDKLDOSEIKSLMCFLYILKNS  780
Qy      1316  DDALFTYNNKASISELMDFTTISEVCLHOFQYWKXVYIARTGMARLQOLGSLDLSLT  1375
Db      781  DDALFTYNNKASISELMDFTTISEVCLHOFQYWKXVYIARTGMARLQOLGSLDLSLT  840
Qy      1376  NNSYGHSDADVHLQSLLEANIATEVCLTALDLSLFTLAFKQQLADHGNPLMKKVPVY  1435
Db      841  NNSYGHSDADVHLQSLLEANIATEVCLTALDLSLFTLAFKQQLADHGNPLMKKVPVY  900
Qy      1436  YLCFLQKQOSETALKVFTALRSLLIYKPSSTFEGRADMCALCYEILKCNKSLSIRT  1495
Db      901  YLCFLQKQOSETALKVFTALRSLLIYKPSSTFEGRADMCALCYEILKCNKSLSIRT  960
Qy      1496  EASQLLYFLMRNPNFDYTGKKSFRTHLYIISVSQILADVIGIGTRFOOSLSIINNCA  1555
Db      961  EASQLLYFLMRNPNFDYTGKKSFRTHLYIISVSQILADVIGIGTRFOOSLSIINNCA  1020
Qy      1556  SDRLIKHTSFSSDVKDLTKRIRTVLMATQMEKEHNDPEMLVDLOYSLKSYASTPELAK  1615
Db      1021  SDRLIKHTSFSSDVKDLTKRIRTVLMATQMEKEHNDPEMLVDLOYSLKSYASTPELAK  1080
Qy      1616  TWLDSMARIVXNGDLSAAMCYVHTALVAEYLTRKEAVONBEPPLPHSHSACLRRSG  1675
Db      1081  TWLDSMARIVXNGDLSAAMCYVHTALVAEYLTRKEAVONBEPPLPHSHSACLRRSG  1140
Qy      1676  GVPROCTAFRYITPNIIDEASMEDVGNQDVHFNEDVIMELLBOCADGLMAERYELIA  1735
Db      1141  GVPROCTAFRYITPNIIDEASMEDVGNQDVHFNEDVIMELLBOCADGLMAERYELIA  1200

```

```

Qy      1736  DIYKLIPIYKRRDPERLALYDTLHRAYSKVTEWMSGRLGLTYFRVAFGQAQOY  1795
Db      1201  DIYKLIPIYKRRDPERLALYDTLHRAYSKVTEWMSGRLGLTYFRVAFGQAQOY  1260
Qy      1796  FTDSSTVEGFEDGKEYIYKEPKLTPLSISQRLKLYSDKGSSENVKMIQDSKVN  1855
Db      1261  FTDSSTVEGFEDGKEYIYKEPKLTPLSISQRLKLYSDKGSSENVKMIQDSKVN  1320
Qy      1856  PKDLDSKAYIQVTHVIPPPEKELQERKTEBERSHNRFFWEMPFTQTKRQGVBEQ  1915
Db      1321  PKDLDSKAYIQVTHVIPPPEKELQERKTEBERSHNRFFWEMPFTQTKRQGVBEQ  1380
Qy      1916  CKRRTILTAHCFPVYKRIIPMYOHTDLNPIEVAIDEMSKVAELRQLCSSAEVDMIX  1975
Db      1381  CKRRTILTAHCFPVYKRIIPMYOHTDLNPIEVAIDEMSKVAELRQLCSSAEVDMIX  1440
Qy      1976  LQKLQGSVSVQVNAAPLAVARAFDDTNTKRYPDNVKYLKEVRFQVEACGQALAVNE  2035
Db      1441  LQKLQGSVSVQVNAAPLAVARAFDDTNTKRYPDNVKYLKEVRFQVEACGQALAVNE  1500
Qy      2036  RLKEDOLEYOEMKANYREMAKELSEIMHBOI  2068
Db      1501  RLKEDOLEYOEMKANYREMAKELSEIMHBOI  1533

RESULT 8
US-09-736-969A-10
; Sequence 10, Application US/09736969A
; Patent No. US2002068302A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-4 Transmembrane Protein
; FILE REFERENCE: 020054-000411US
; CURRENT APPLICATION NUMBER: US/09/736,969A
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1534

```

TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: KIAA1058  
 US-09-736-969A-10

Query Match 72.9%; Score 7976; DB 10; Length 1534;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 536 ASGNLDKNAKFAAIYQDSNKLNDMLKLLADFKRPEKAKLPVILGMLDITIDNVSD 595  
 DB 1 ASGNLDKNAKFAAIYQDSNKLNDMLKLLADFKRPEKAKLPVILGMLDITIDNVSD 60  
 QY 596 PNNYNSYIPTKOFETCSKPTTFEEVEEFCIPKHTQFYITITNHLVYPRYLYKDYQ 655  
 DB 61 PNNYNSYIPTKOFETCSKPTTFEEVEEFCIPKHTQFYITITNHLVYPRYLYKDYQ 120  
 QY 656 KAPAKARNIAICIEFKDSEEDSQPLKCIYGRPGVFTRISAPAAVLHHQNPFEYDEIK 715  
 DB 121 KAPAKARNIAICIEFKDSEEDSQPLKCIYGRPGVFTRISAPAAVLHHQNPFEYDEIK 180  
 QY 716 IELPTQLHKKHLLTFHFVSCDNSKSTKKRDVETQVGYSMPLKLDGRVYVSEQHI 775  
 DB 181 IELPTQLHKKHLLTFHFVSCDNSKSTKKRDVETQVGYSMPLKLDGRVYVSEQHI 240  
 QY 776 PVSANLPSGYLGQBLGMRHYPGLIKWVDGKPLKISTHLVSTVYVYVYVYVYVYVYVYV 835  
 DB 241 PVSANLPSGYLGQBLGMRHYPGLIKWVDGKPLKISTHLVSTVYVYVYVYVYVYVYVYV 300  
 QY 836 QKTESGAQALGNELVYKLSLHAMEGHVIAFLPTILNQLFRVLTATQGEVAVVNTVYI 895  
 DB 301 QKTESGAQALGNELVYKLSLHAMEGHVIAFLPTILNQLFRVLTATQGEVAVVNTVYI 360  
 QY 896 IHVVAQCHEEGESHRSYVYKAYKAEPYVASRYKTVHEELTKSMITLIKESADFLTSNK 955  
 DB 361 IHVVAQCHEEGESHRSYVYKAYKAEPYVASRYKTVHEELTKSMITLIKESADFLTSNK 420  
 QY 956 LTKYSWFFPDVLIKMAQHLIENSKYKCLRNOGFPASVYHAYETVVMMLPHITQKFRN 1015  
 DB 421 LTKYSWFFPDVLIKMAQHLIENSKYKCLRNOGFPASVYHAYETVVMMLPHITQKFRN 480  
 QY 1016 PEASKNANSLAVFIKRCFTFMDRGVFVKQINNYISCFAPGDEKTLFEYKFEFLRVVCM 1075  
 DB 481 PEASKNANSLAVFIKRCFTFMDRGVFVKQINNYISCFAPGDEKTLFEYKFEFLRVVCM 540  
 QY 1076 EHYIPLANLPMPPGKRIQRYODQLDYSLTDFECRNHFIYGLILRVEGTALQDFREVL 1135  
 DB 541 EHYIPLANLPMPPGKRIQRYODQLDYSLTDFECRNHFIYGLILRVEGTALQDFREVL 600  
 QY 1136 AISVLKNLIIKSPDDRVSRSRSHQARIATLYLPLFGLILENORINVRVSPPVVAGMT 1195  
 DB 601 AISVLKNLIIKSPDDRVSRSRSHQARIATLYLPLFGLILENORINVRVSPPVVAGMT 660  
 QY 1196 VKDESIALPVPVPLVTPQKSTLDSNLHDLGALSIGLASPYTSTPININSVANDSRGS 1255  
 DB 661 VKDESIALPVPVPLVTPQKSTLDSNLHDLGALSIGLASPYTSTPININSVANDSRGS 720  
 QY 1256 L1STDSGNSLPERNSKSNLDRKQSSSTLGNVVRCDKLDQSEIKSLMCFYIYIKSN 1315  
 DB 721 L1STDSGNSLPERNSKSNLDRKQSSSTLGNVVRCDKLDQSEIKSLMCFYIYIKSN 780  
 QY 1316 DDALFTYNNKASTSELMDFITISEVCLHOFVWGXKXIARTGMHARLQOLSLNSLTF 1375  
 DB 781 DDALFTYNNKASTSELMDFITISEVCLHOFVWGXKXIARTGMHARLQOLSLNSLTF 840  
 QY 1376 NNSYGHSDADVHLQSLLEANIATEVCLTALDLSLFTLAFAKQOLADHGNPLMKVFPVY 1435  
 DB 841 NNSYGHSDADVHLQSLLEANIATEVCLTALDLSLFTLAFAKQOLADHGNPLMKVFPVY 900  
 QY 1436 YICFLQKQSEETALKVFTALNSLIYKFPSTFEGBAIDMCALCYEILKCCNSKLSIRT 1495  
 DB 901 YICFLQKQSEETALKVFTALNSLIYKFPSTFEGBAIDMCALCYEILKCCNSKLSIRT 960

QY 1496 EASQLLYFLMRNNEPDYTKSKSFVRTHLOVIVISVQLIADVVGIGTRQOQSIINNCA 1555  
 DB 961 EASQLLYFLMRNNEPDYTKSKSFVRTHLOVIVISVQLIADVVGIGTRQOQSIINNCA 1020  
 QY 1556 SDRILKHTSPSSDVKDLTKRIRTVLMATAQKHEHNDPEMLVDLOYSLAKSYASTPELRK 1615  
 DB 1021 SDRILKHTSPSSDVKDLTKRIRTVLMATAQKHEHNDPEMLVDLOYSLAKSYASTPELRK 1080  
 QY 1616 TMLDSMAIHYNNGDLSAAMCYVHTALVVEYLTRKSAVOMEPPLPHSHACLRBSRG 1675  
 DB 1081 TMLDSMAIHYNNGDLSAAMCYVHTALVVEYLTRKSAVOMEPPLPHSHACLRBSRG 1140  
 QY 1676 GVFRQCTAFRYITPNIDEASMMEDVGMQVDFHFNEDVLMELBQCADGLMAERYEYELIA 1735  
 DB 1141 GVFRQCTAFRYITPNIDEASMMEDVGMQVDFHFNEDVLMELBQCADGLMAERYEYELIA 1200  
 QY 1736 DIYKLIIPYERKRDPERLAHLVDTLHRAYSKTEVMSGRLLCTYFRVAFGQAQYQ 1795  
 DB 1201 DIYKLIIPYERKRDPERLAHLVDTLHRAYSKTEVMSGRLLCTYFRVAFGQAQYQ 1260  
 QY 1796 FTDSETEVGEFDEDEGKEYIYKEPKLTPLEISORLLKYSDKGSNNVMQIDSGKN 1855  
 DB 1261 FTDSETEVGEFDEDEGKEYIYKEPKLTPLEISORLLKYSDKGSNNVMQIDSGKN 1320  
 QY 1856 PKDLSKTAAYIQVTHVIFPDEKELQERKTEPERSHNIRFMEFPTQYGRQGVBEQ 1915  
 DB 1321 PKDLSKTAAYIQVTHVIFPDEKELQERKTEPERSHNIRFMEFPTQYGRQGVBEQ 1380  
 QY 1916 CKRRITLAIHCFPYVKRIPVMYQHTNDLPIEVALIDEMSKVALEQLCGSAEVDIMK 1975  
 DB 1381 CKRRITLAIHCFPYVKRIPVMYQHTNDLPIEVALIDEMSKVALEQLCGSAEVDIMK 1440  
 QY 1976 LQKIQGSVSQVANGPLAVARAFLDPTNTRXPBNKTKLKEVROVPEAGQALVNE 2035  
 DB 1441 LQKIQGSVSQVANGPLAVARAFLDPTNTRXPBNKTKLKEVROVPEAGQALVNE 1500  
 QY 2036 RLKEDQLEYOEBEMKANYREMAKELSEIMHEQI 2068  
 DB 1501 RLKEDQLEYOEBEMKANYREMAKELSEIMHEQI 1533

RESULT 9  
 US-09-736-960-10  
 ; Sequence 10, Application US/09736960  
 ; Patent No. US20020102267A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lu, Peter  
 ; APPLICANT: Garman, Jonathan David  
 ; APPLICANT: Candia III, Albert Frederick  
 ; APPLICANT: Arbor Vita Corporation  
 ; TITLE OF INVENTION: CLASP-5 Transmembrane Protein  
 ; FILE REFERENCE: 020054-000511US  
 ; CURRENT APPLICATION NUMBER: US/09/736, 960  
 ; PRIOR FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: US 60/160, 860  
 ; PRIOR FILING DATE: 1999-10-21  
 ; PRIOR APPLICATION NUMBER: US 60/162, 498  
 ; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: US 60/170, 453  
 ; PRIOR FILING DATE: 1999-12-13  
 ; PRIOR APPLICATION NUMBER: US 60/176, 195  
 ; PRIOR FILING DATE: 2000-01-14  
 ; PRIOR APPLICATION NUMBER: US 60/182, 296  
 ; PRIOR FILING DATE: 2000-02-14  
 ; PRIOR APPLICATION NUMBER: US 09/547, 276  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196, 267  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196, 460  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196, 527  
 ; PRIOR FILING DATE: 2000-04-11





; PRIOR APPLICATION NUMBER: US 60/162,498  
 ; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: US 60/170,453  
 ; PRIOR FILING DATE: 1999-12-13  
 ; PRIOR APPLICATION NUMBER: US 60/176,195  
 ; PRIOR FILING DATE: 2000-01-14  
 ; PRIOR APPLICATION NUMBER: US 60/182,296  
 ; PRIOR FILING DATE: 2000-02-14  
 ; PRIOR APPLICATION NUMBER: US 09/547,276  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196,267  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196,460  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196,527  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 60/196,528  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: US 09/687,837  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/240,503  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/240,508  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/240,539  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/240,543  
 ; PRIOR FILING DATE: 2000-10-13  
 ; NUMBER OF SEQ ID NOS: 115  
 ; SOFTWARE: Patent Ver. 2.1  
 ; SEQ ID NO: 9  
 ; LENGTH: 1352  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: human CLASP-2A  
 ; US-09-736-968A-9

Query Match 62.0%; Score 6784.5; DB 9; Length 1352;  
 Best Local Similarity 92.7%; Pred. No. 0;  
 Matches 1355; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLNHHNPFYDEIKIELPTOLHEKHLITTFHVSCDSKSGSTKKRDVETOVGYSL 760  
 DB 1 VLNHHNPFYDEIKIELPTOLHEKHLITTFHVSCDSKSGSTKKRDVETOVGYSL 60  
 QY 761 PLLKGRVVTSECHITVSAANLPSGYGYOLGMRHYGPEIKWVDSGKFLKISTHLVST 820  
 DB 61 PLLKGRVVTSECHITVSAANLPSGYGYOLGMRHYGPEIKWVDSGKFLKISTHLVST 120  
 QY 821 VTTODHNLNPFYOCCOTSSGAOLGNELVKYLKSLHAMEGHVMTAFLPTIINOLFRVLT 880  
 DB 121 VTTODHNLNPFYOCCOTSSGAOLGNELVKYLKSLHAMEGHVMTAFLPTIINOLFRVLT 180  
 QY 881 RATOBEVANVTRVIIHVVAQCHBEGLESLSYVYKA/YKAEYVASEYKTVHEBELTKSM 940  
 DB 181 RATOBEVANVTRVIIHVVAQCHBEGLESLSYVYKA/YKAEYVASEYKTVHEBELTKSM 240  
 QY 941 TTIILKRSAPLTSNKLKTSWFFFDVLTISMAOHLIENSKVTLNORFPASTYHAAVETV 1000  
 DB 241 TTIILKRSAPLTSNKLKTSWFFFDVLTISMAOHLIENSKVTLNORFPASTYHAAVETV 300  
 QY 1001 VNMMLPHITOKFRDNEPASKNANHSIAVFIKRCFTFMDRGPFVKOINNTISCAPDDEPT 1060  
 DB 301 VNMMLPHITOKFRDNEPASKNANHSIAVFIKRCFTFMDRGPFVKOINNTISCAPDDEPT 360  
 QY 1061 LFEYKEFLRVVNCNHEHYIPLNLPMPFGKRIQRYODLQLDVSLTDFCRNHFVGLILR 1120  
 DB 361 LFEYKEFLRVVNCNHEHYIPLNLPMPFGKRIQRYODLQLDVSLTDFCRNHFVGLILR 420  
 QY 1121 EVGTALQERREVRLLAISVLKULLIGSDDRYASGSHQARIATLYLPFGILLIENVQRI 1180  
 DB 421 EVGTALQERREVRLLAISVLKULLIGSDDRYASGSHQARIATLYLPFGILLIENVQRI 480

QY 1181 NVRDVSPPPVNAGMTVKDESIALPAVNPLVTPQKSTLDNSLHKDLLAIGSIASPYTTS 1240  
 DB 481 NVRDVSPPPVNAGMTVKDESIALPAVNPLVTPQKSTLDNSLHKDLLAIGSIASPYTTS 540  
 QY 1241 TPININSVANADSRGSLISTDSGNSLPERNSEKNSLDHQOOSTLGNSVVRCDKLDQSEI 1300  
 DB 541 TPININSVANADSRGSLISTDSGNSLPERNSEKNSLDHQOOSTLGNSVVRCDKLDQSEI 600  
 QY 1301 KSLMCFIYIILKSNBDDLFYWNKASTSELMDFPITSEVCLHQOYGVKXYIA----- 1354  
 DB 601 KSLMCFIYIILKSNBDDLFYWNKASTSELMDFPITSEVCLHQOYGVKXYIA----- 1354  
 QY 1355 -----RTGMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLLEANIA 1397  
 DB 661 GPVYHDKSQTLPIVGRNRTGMHARLQOLGSLDNLSTFNHSGHSDADVLHOSLLEANIA 720  
 QY 1398 TEVCLTALDLSLFTLAFKNOLLADHGNPLMKVFDVYLCPLOKHGETALKNVFTALR 1457  
 DB 721 TEVCLTALDLSLFTLAFKNOLLADHGNPLMKVFDVYLCPLOKHGETALKNVFTALR 780  
 QY 1458 SLIYKPSSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLYFLARNNFDYTKGKSF 1517  
 DB 781 SLIYKPSSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLYFLARNNFDYTKGKSF 840  
 QY 1518 VRTHLQVLIISVSQLADVVGIGTRFOOSLSTINNCASDRLIKRTSFSSPDVCDLTKRIR 1577  
 DB 841 VRTHLQVLIISVSQLADVVGIGTRFOOSLSTINNCASDRLIKRTSFSSPDVCDLTKRIR 900  
 QY 1578 TVLMATAQMKHENDPEMLVDLQVSLAKSVASTPELRTKWLDSMARIHVKNQDLSBAAMC 1637  
 DB 901 TVLMATAQMKHENDPEMLVDLQVSLAKSVASTPELRTKWLDSMARIHVKNQDLSBAAMC 960  
 QY 1638 YVHTALVAEYLTTRKANQVMBPRLPHSHSACLRSGVRRQGTARVITPNIDERS 1697  
 DB 961 YVHTALVAEYLTTRK-----GVFRQGTARVITPNIDERS 997  
 QY 1698 NMEDVGMQDVHFNEDVLMELBOCADGMLKABERYELADIYKLIPIYEKRDFFRLAHL 1757  
 DB 998 NMEDVGMQDVHFNEDVLMELBOCADGMLKABERYELADIYKLIPIYEKRDFFRLAHL 1051  
 QY 1758 YDTLIRAYSKYTEVNSGRRLGTYFRVAFGQAQOYFTDSETDVEGFEDDEGKEYIY 1817  
 DB 1052 -----FEDEDEGKEYIY 1062  
 QY 1818 KEPKLTPISEISQRLKYSDFGSENVKMTODSKNVPKOLDSKTAIYOTHTVIPPDE 1877  
 DB 1063 KEPKLTPISEISQRLKYSDFGSENVKMTODSKNVPKOLDSKTAIYOTHTVIPPDE 1122  
 QY 1878 KELDERKTEPFRSHNIRBFMEPMPPTONGKRGOGVEGCKRRTITLAIHCPYVKGRIY 1937  
 DB 1123 KELDERKTEPFRSHNIRBFMEPMPPTONGKRGOGVEGCKRRTITLAIHCPYVKGRIY 1182  
 QY 1938 MYOHTDNLPIEVALIDEMSKVABLRQLCSSAEVDMILQULQGSVSVQVNAAGLAYAR 1997  
 DB 1183 MYOHTDNLPIEVALIDEMSKVABLRQLCSSAEVDMILQULQGSVSVQVNAAGLAYAR 1242  
 QY 1998 AFLDPTNTKRIYDNDKVKLLKXVFPFVACGQALAVNERLKEDELEVOEBKANYREMA 2057  
 DB 1243 AFLDPTNTKRIYDNDKVKLLKXVFPFVACGQALAVNERLKEDELEVOEBKANYREMA 1302  
 QY 2058 KELSEIMHEOICPLEKTSVLPNSLHIFNAISGPTSTMVGMTSSSSSV 2107  
 DB 1303 KELSEIMHEOICPLEKTSVLPNSLHIFNAISGPTSTMVGMTSSSSSV 1352

RESULT 11  
 US-09-736-969A-9  
 ; Sequence 9, Application US/09736969A  
 ; Patent No. US20020068302A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lu, Peter  
 ; APPLICANT: Garman, Jonathan David

APPLICANT: Candia III, Albert Frederick  
APPLICANT: Arbor Vita Corporation  
TITLE OF INVENTION: CLASP-4 Transmembrane Protein  
FILE REFERENCE: 020054-00041US  
CURRENT APPLICATION NUMBER: US/09/736,969A  
CURRENT FILING DATE: 2000-12-13  
PRIOR APPLICATION NUMBER: US 60/160,860  
PRIOR FILING DATE: 1999-10-21  
PRIOR APPLICATION NUMBER: US 60/162,498  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 60/170,453  
PRIOR FILING DATE: 1999-12-13  
PRIOR APPLICATION NUMBER: US 60/176,195  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: US 60/182,296  
PRIOR FILING DATE: 2000-02-14  
PRIOR APPLICATION NUMBER: US 09/547,276  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,267  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,460  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,527  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,528  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 09/687,837  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,503  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,508  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,539  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,543  
PRIOR FILING DATE: 2000-10-13  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 1352  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human CLASP-2  
US-09-736-969A-9

Query Match 62.0%; Score 6784.5; DB 10; Length 1352;  
Best Local Similarity 92.7%; Pred. No. 0;  
Matches 1345; Conservative 1; Mismatches 3; Indels 101; Gaps 3;  
QY 701 VLIHHNPFYDIKIELPTQLHEKHLITLFFHVS CDNSKSGSTKKRUVETQVGSML 760  
DB 1 VLIHHNPFYDIKIELPTQLHEKHLITLFFHVS CDNSKSGSTKKRUVETQVGSML 60  
QY 761 PLIKDGRVVTSEGIHIVSANLPBGYLGYOELGWRHYGPEIKWVDGKPLKISTHLVST 820  
DB 61 PLIKDGRVVTSEGIHIVSANLPBGYLGYOELGWRHYGPEIKWVDGKPLKISTHLVST 120  
QY 821 VTTODHNLNFPYCKTSSGAOLGNELVYKLSIHAHEGHWIAFLPTLIINQLFRVLT 880  
DB 121 VTTODHNLNFPYCKTSSGAOLGNELVYKLSIHAHEGHWIAFLPTLIINQLFRVLT 180  
QY 881 RATOEEVAVNVTRVLIHVAQAQCHEGLESLSYVAYKAPRYASEKTYHSELTKSM 940  
DB 181 RATOEEVAVNVTRVLIHVAQAQCHEGLESLSYVAYKAPRYASEKTYHSELTKSM 240  
QY 941 TTIILKSADELTSNKLKTSWFFDVLIKSMQHLIENSKVKLLRNQRPASYHAAVETV 1000  
DB 241 TTIILKSADELTSNKLKTSWFFDVLIKSMQHLIENSKVKLLRNQRPASYHAAVETV 300  
QY 1001 VNMMLHITQKFRDNEASKNANHSIAVTKRCFTTMDRGFVFKQIINNYISCFAPGDPKT 1060  
DB 301 VNMMLHITQKFRDNEASKNANHSIAVTKRCFTTMDRGFVFKQIINNYISCFAPGDPKT 360

QY 1061 LPEYKFEFLRVYCNHEHYIPLNLPMPFGKRIQRYQDQLDYSLDECRNHFVGLLRL 1120  
DB 361 LPEYKFEFLRVYCNHEHYIPLNLPMPFGKRIQRYQDQLDYSLDECRNHFVGLLRL 420  
QY 1121 EVGTALQEFREVRLLIIVLKNLLIKHSFDDRYSRSHQARIATLYLPLFGLLIENVORI 1180  
DB 421 EVGTALQEFREVRLLIIVLKNLLIKHSFDDRYSRSHQARIATLYLPLFGLLIENVORI 480  
QY 1181 NVRDVSPPPVNAQGMVVKDESLAPVNPVLNPKQSTLDNSLHKDLGALIGISAPYTS 1240  
DB 481 NVRDVSPPPVNAQGMVVKDESLAPVNPVLNPKQSTLDNSLHKDLGALIGISAPYTS 540  
QY 1241 TPINISVNNADSRGSLISTDSGNSLPERNSEKNSLDKHOQSTLGNSVVRCDKQDSEI 1300  
DB 541 TPINISVNNADSRGSLISTDSGNSLPERNSEKNSLDKHOQSTLGNSVVRCDKQDSEI 600  
QY 1301 KSLMCFYIILKSMDDALFTYWNKASTSEIMDFTTISEVCLHOFQYNGKRYIA----- 1354  
DB 601 KSLMCFYIILKSMDDALFTYWNKASTSEIMDFTTISEVCLHOFQYNGKRYIA----- 660  
QY 1355 -----RTGMMHARLQOLGSLDNSLTFNHSGHSDADVLRHOSLEANAIA 1397  
DB 661 GPIVHDRKSQTLPVSRNRTGMMHARLQOLGSLDNSLTFNHSGHSDADVLRHOSLEANAIA 720  
QY 1398 TEVCCLTALDTLSLFTLAFKNQLADHGNPMLKQVFDVYLGFLOHGOSETALKVFTALR 1457  
DB 721 TEVCCLTALDTLSLFTLAFKNQLADHGNPMLKQVFDVYLGFLOHGOSETALKVFTALR 780  
QY 1458 SLIYFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLYFLMRNPNFDYTGKXSF 1517  
DB 781 SLIYFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLYFLMRNPNFDYTGKXSF 840  
QY 1518 VRTHLQVLIISVQLADVVGIGTFEFOQSLIINNCAUSDRLIKRTSSSDVKDLTKRIR 1577  
DB 841 VRTHLQVLIISVQLADVVGIGTFEFOQSLIINNCAUSDRLIKRTSSSDVKDLTKRIR 900  
QY 1578 TVLMATAQMKHEHNPENLVLDQYSLAKSYASTPELRATWLDMSMARIHVKNQDLSSEAMC 1637  
DB 901 TVLMATAQMKHEHNPENLVLDQYSLAKSYASTPELRATWLDMSMARIHVKNQDLSSEAMC 960  
QY 1638 YVHTALVAEYLTRKEAVQWEPPLPHSHSACLRSSRGVFRQGTARVITPNIDEAS 1697  
DB 961 YVHTALVAEYLTRK-----GVFRQGTARVITPNIDEAS 997  
QY 1698 NMEDVGMQDVHFNEDVLMELLBOCADLWKAERYELADIYKLIIPYEKKRDEPRLAHL 1757  
DB 998 NMEDVGMQDVHFNEDVLMELLBOCADLWKAERYELADIYKLIIPYEKKRDEPRLAHL 1051  
QY 1758 YDTLRAVSKYEVNHSGRRLIGTFYFVAFGQAQYOFTDSEITDVEGFFDEDEKEXIY 1817  
DB 1052 -----FDEDEKEXIY 1062  
QY 1818 KEPKLTPLSEISQRLIKLYSDKFGSENVKMIQDSKVPKOLDSKAYAVIQTTHVIPFDE 1877  
DB 1063 KEPKLTPLSEISQRLIKLYSDKFGSENVKMIQDSKVPKOLDSKAYAVIQTTHVIPFDE 1122  
QY 1878 KELORKEPERSHNRIRFEMFEMPTONGKQOGGVEOCCRRITLTAHCPFPYVKRIRPV 1937  
DB 1123 KELORKEPERSHNRIRFEMFEMPTONGKQOGGVEOCCRRITLTAHCPFPYVKRIRPV 1182  
QY 1938 MYOHTDLNPIEVALIDEMSKYAEELQJCSAEVDMIKLOKLQGSVSVQVNAAGLAAAR 1997  
DB 1183 MYOHTDLNPIEVALIDEMSKYAEELQJCSAEVDMIKLOKLQGSVSVQVNAAGLAAAR 1242  
QY 1998 AFLDNTNKRYPDNKVKLLKEVFRQFVACGQALAVNERLIKEDQLEYOEEMKANRYREMA 2057  
DB 1243 AFLDNTNKRYPDNKVKLLKEVFRQFVACGQALAVNERLIKEDQLEYOEEMKANRYREMA 1302  
QY 2058 KELSIMEOICPLEBKSYSVPLNSLIHFNALISGPTSTNMHGMSSSSSVV 2107  
DB 1303 KELSIMEOICPLEBKSYSVPLNSLIHFNALISGPTSTNMHGMSSSSSVV 1352

RESULT 12  
US-09-736-960-9  
Sequence 9, Application US/09736960  
Patent No. US20020102267A1  
GENERAL INFORMATION:  
APPLICANT: Lu, Peter  
APPLICANT: Gorman, Jonathan David  
APPLICANT: Candia III, Albert Frederick  
APPLICANT: Arbor Vita Corporation  
TITLE OF INVENTION: CLASP-5 Transmembrane Protein  
FILE REFERENCE: 020054-0005105  
CURRENT APPLICATION NUMBER: US/09/736,960  
PRIOR FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: US 60/160,860  
PRIOR FILING DATE: 1999-10-21  
PRIOR APPLICATION NUMBER: US 60/162,498  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 60/170,453  
PRIOR FILING DATE: 1999-12-13  
PRIOR APPLICATION NUMBER: US 60/176,195  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: US 60/182,296  
PRIOR FILING DATE: 2000-02-14  
PRIOR APPLICATION NUMBER: US 09/547,276  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,267  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,460  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,527  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 60/196,528  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 09/687,837  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,503  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,508  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,539  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/240,543  
PRIOR FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 134  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 9  
LENGTH: 1352  
TYPE: PR1  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human CLASP-2A  
US-09-736-960-9  
Query Match 62.0%; Score 6784.5; DB 10; Length 1352;  
Best Local Similarity 92.7%; Pred. No. 0;  
Matches 135; Conservative 1; Mismatches 3; Indels 101; Gaps 3;  
QY 701 VHHHNNPEFYDEIKELPTQLHEKHHLLTFPHVSCDSSKSTKRDVETOVGYSWL 760  
DB 1 VHHHNNPEFYDEIKELPTQLHEKHHLLTFPHVSCDSSKSTKRDVETOVGYSWL 60  
QY 761 PLLKGRVVTSEHPIVSNLPSGYLGQELGMRHYGPRIKWDGKPLKISTHLVST 820  
DB 61 PLLKGRVVTSEHPIVSNLPSGYLGQELGMRHYGPRIKWDGKPLKISTHLVST 120  
QY 821 VTTODQHLNFPQYCKCTESGAQALGNELVKYLSLHAMEGHVMIAPLFTIINQLFRVLT 880  
DB 121 VTTODQHLNFPQYCKCTESGAQALGNELVKYLSLHAMEGHVMIAPLFTIINQLFRVLT 180  
QY 881 RATOEVAVNVTRVIIHVVAQCHEBGLSHLSYKYAKAEPPYVASEYKTVHEELTKSM 940  
DB 181 RATOEVAVNVTRVIIHVVAQCHEBGLSHLSYKYAKAEPPYVASEYKTVHEELTKSM 240

QY 941 TTIKPSADFLTSNKLKYSWFFPDVLIKSAQHLIENSKVLLJNQHPPASVYHNAVETV 1000  
DB 241 TTIKPSADFLTSNKLKYSWFFPDVLIKSAQHLIENSKVLLJNQHPPASVYHNAVETV 300  
QY 1001 VNMILPHITQKFRDNPASKNANSLAVFIRCFETMDRGVYFKQINNYISCPAGDPKT 1060  
DB 301 VNMILPHITQKFRDNPASKNANSLAVFIRCFETMDRGVYFKQINNYISCPAGDPKT 360  
QY 1061 LPEYKFEELRVACNHEHYIPLNMPFKSGRIQRQDQLDLSLDEFCRNHFLVGLLR 1120  
DB 361 LPEYKFEELRVACNHEHYIPLNMPFKSGRIQRQDQLDLSLDEFCRNHFLVGLLR 420  
QY 1121 EYGTALQFREVRLAISVLKNLILKHSFDDRVSRSQARATATYLPFGILLINVORI 1180  
DB 421 EYGTALQFREVRLAISVLKNLILKHSFDDRVSRSQARATATYLPFGILLINVORI 480  
QY 1181 NVRDVSPPPVNAQMTVKESLALPAVNPLVTPQKSTLDNSLHKDLGASIGIAPYTT 1240  
DB 481 NVRDVSPPPVNAQMTVKESLALPAVNPLVTPQKSTLDNSLHKDLGASIGIAPYTT 540  
QY 1241 TPINISVNNADSRGSLISTDGSNSLPERNSEKNSLDKHOOSTTGNSVVRCDKDOSEI 1300  
DB 541 TPINISVNNADSRGSLISTDGSNSLPERNSEKNSLDKHOOSTTGNSVVRCDKDOSEI 600  
QY 1301 KSLMCPYIILKSMDDALFTYWNKASTSELMDFETISEVCLHOFQYMGKRYIA----- 1354  
DB 601 KSLMCPYIILKSMDDALFTYWNKASTSELMDFETISEVCLHOFQYMGKRYIA----- 660  
QY 1355 -----RTGMMHARLQGLSDNSLTENHSYGSADADVLHQSILEANIA 1397  
DB 661 GPIVHDRKSQTLPVSRNRTGMMHARLQGLSDNSLTENHSYGSADADVLHQSILEANIA 720  
QY 1398 TEVCITLADTSLFTLAKNQLADHGNPLMKKFPDYVLCLOHQGSETALKNFTLYR 1457  
DB 721 TEVCITLADTSLFTLAKNQLADHGNPLMKKFPDYVLCLOHQGSETALKNFTLYR 780  
QY 1458 SLIYKFPSTFEGRADMCALCYELIKCCNSKLSIRTEASQLLYLMRNPNFDYTKSFP 1517  
DB 781 SLIYKFPSTFEGRADMCALCYELIKCCNSKLSIRTEASQLLYLMRNPNFDYTKSFP 840  
QY 1518 VATHLOVITISVQLIADVVGIGTGFQOGLSIINNCA NSDRILIKHTSFSSDVYDKLRIR 1577  
DB 841 VATHLOVITISVQLIADVVGIGTGFQOGLSIINNCA NSDRILIKHTSFSSDVYDKLRIR 900  
QY 1578 TYLMATAQMKHEHNPENLVLDQYSLAGSYASTPELRKTYLDSMARIHVKNQDLSBANMC 1637  
DB 901 TYLMATAQMKHEHNPENLVLDQYSLAGSYASTPELRKTYLDSMARIHVKNQDLSBANMC 960  
QY 1638 YVHTVALVAEYLTRKEAVQWMBPRLPHSHSACLRSGRGVFPQGCYAFRVITPNIIDEAS 1697  
DB 961 YVHTVALVAEYLTRK-----GVRRQGCYAFRVITPNIIDEAS 997  
QY 1698 NMEDVGMQDVHNEEDVNLLEQCADGLMKERYELIADYKLIIPYKERDFERLAL 1757  
DB 998 NMEDVGMQDVHNEEDVNLLEQCADGLMKERYELIADYKLIIPYKERDF----- 1051  
QY 1758 YDTLRAVSKTYEVMHSGRLLGTYFRVAFPGQAQOYPTDSETVYEGFPEDEGKEYTY 1817  
DB 1052 -----FEDEDEKEYTY 1062  
QY 1818 KEPKLTPLSEISQRLKLKXSPKGSENVKMTODSGKNPKOLDSKYAYIQVHVLPFDE 1877  
DB 1063 KEPKLTPLSEISQRLKLKXSPKGSENVKMTODSGKNPKOLDSKYAYIQVHVLPFDE 1122  
QY 1878 KELQERKTEPERSHNIRPFMEPEPTQYKQGGVYEOCKRTIITAIHCFYVKKRIPV 1937  
DB 1123 KELQERKTEPERSHNIRPFMEPEPTQYKQGGVYEOCKRTIITAIHCFYVKKRIPV 1182  
QY 1938 MYQHTDLNPIEVALIDEMSKYAEIRQICSSAEVMIKQLKQSSVSVQVNAAGPLAYAR 1997  
DB 1183 MYQHTDLNPIEVALIDEMSKYAEIRQICSSAEVMIKQLKQSSVSVQVNAAGPLAYAR 1242



Db 1284 MCYLIVKVISSEDTLLTYNNKVSPOELINILVLEVCLEPHRYMGKRNIFARVHDAWLSKH 1343  
Qy 1355 -----RTGMHARLOOLGSLDNLFFNHSYSHSDADVLHQSLEANIATE 1399  
Db 1344 FGDIDKSCQMPALRRRSVGMARLQHSLESSFTLNHSSATTEADIFQALLEGNTATE 1403  
Qy 1400 VCLTALDLSLFTLAKQQLADHGNPLMKYFVYVLCFLQKHOSETALKNVFTALRSL 1459  
Db 1404 VSLTLETSIFPTQCFKQQLNNDGNPLMKVFDIHLAFLANGOSEVSLKHVPSLRSF 1463  
Qy 1460 IYKPFSTFEGRADMCALCYELKCNKSLSIETBASQLLYFLMRNFDYTGKKSFR 1519  
Db 1464 ISKPSAFKFGKVNCAFCYVLCCKTSKISSTRBASALYLLMRNFEYTKRTFAR 1523  
Qy 1520 THLOVYISVOLIAVVGIGSTRFOOSLIINNANSDDLHHTSSPSVQLTRIRIV 1579  
Db 1524 THLOVYISVOLIAVVGIGSTRFOOSLIINNANSDDLHHTSSPSVQLTRIRIV 1583  
Qy 1580 LMATAQMKHENDPEMLVDLOYSLAKSYASTPELRKTMLSMARLHVKNGLSEAMCY 1639  
Db 1584 LMATAQMKHENDPEMLVDLOYSLAKSYASTPELRKTMLSMARLHVKNGLSEAMCY 1643  
Qy 1640 HTTALVAELTRKEAVQWERPLLPHSHSACLRSSRGCVFRQCTAFRVTINIDBASMM 1699  
Db 1644 HVAALVAELTRKEAVQWERPLLPHSHSACLRSSRGCVFRQCTAFRVTINIDBASMM 1680  
Qy 1700 EDVGNODVHFNEDVLMELLEGADGIMKAREYELIADYICLIPIYERKRDPERLAHLVD 1759  
Db 1681 EDAGMDVAVSEVLELEQCVDGIMKAREYEVSEISGLIPIYERKREKLTQYVR 1740  
Qy 1760 TLHRAVSKVTEVMSHSGRLLGTYFRVAFPGQAQOFTDSEZDVSGFFEDSGEKIYKE 1819  
Db 1741 TLHRAVSKVTEVMSHSGRLLGTYFRVAFPGQAQOFTDSEZDVSGFFEDSGEKIYKE 1786  
Qy 1820 PULTLSEISQRLKLYSDKFSSENVKMTODSGKVNPKDLSKAVIQTTHVIPPDEKE 1879  
Db 1787 PULTLSEISQRLKLYSDKFSSENVKMTODSGKVNPKDLSKAVIQTTHVIPPDEKE 1846  
Qy 1880 LOERTPEFRSHNIRFMEEMPTOTGKQSGVSEBQCKRTILTMHCPYKRIIPVNY 1939  
Db 1847 LOERTPEFRSHNIRFMEEMPTOTGKQSGVSEBQCKRTILTMHCPYKRIIPVNY 1906  
Qy 1940 QHHTDLNPLEVALIDEMSKVAELRQCSSAEYDMIKLOXLQSGSVQVNAAGPLAVARAF 1999  
Db 1907 EGVVNLKPIDVATDEIKDCTALHLKCSSVDVDMLOLQKLOGCYSVQVNAAGPLAVARAF 1966  
Qy 2000 LDDTNTKRYPDNKVKLKEVFRQVYACGOALAVNERLIKEDQLEYQEBMKANYREMAKE 2059  
Db 1967 LNESQANKYPRPKVNELKDMFRKFIQACISALELNERLIKEDQLEYHGLKSNPRDMYKE 2026  
Qy 2060 LSEIMHEQICPLEKTSVLP---NSLHTNALSIPPTS 2094  
Db 2027 LSDIHEQI--LQEDTMHSPMNNLTLLHVCALISGTSN 2062

RESULT 14  
US-09-736-968A-105  
; Sequence 105, Application US/09736968A  
; Patent No. US20020169283A1  
; GENERAL INFORMATION:  
; APPLICANT: Lu, Peter  
; APPLICANT: Garman, Jonathan David  
; APPLICANT: Candia III, Albert Frederick  
; APPLICANT: Arbor Vita Corporation  
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein  
; FILE REFERENCE: 020054-000611US  
; CURRENT APPLICATION NUMBER: US/09/736,968A  
; CURRENT FILING DATE: 2000-12-13  
; PRIOR APPLICATION NUMBER: US 60/160,860  
; PRIOR FILING DATE: 1999-10-21  
; PRIOR APPLICATION NUMBER: US 60/162,498  
; PRIOR FILING DATE: 1999-10-29

; PRIOR APPLICATION NUMBER: US 60/170,453  
; PRIOR FILING DATE: 1999-12-13  
; PRIOR APPLICATION NUMBER: US 60/176,195  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: US 60/182,296  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: US 09/547,276  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: US 60/196,267  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: US 60/196,460  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: US 60/196,527  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: US 60/196,528  
; PRIOR FILING DATE: 2000-04-11  
; PRIOR APPLICATION NUMBER: US 09/687,837  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/240,503  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/240,508  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/240,539  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/240,543  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 115  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 105  
; LENGTH: 2008  
; TYPE: PR  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human CLASP-4  
US-09-736-968A-105

Query Match 55.8%; Score 6105.5; DB 9; Length 2008;  
Best Local Similarity 57.8%; Pred. No. 0;  
Matches 1197; Conservative 327; Mismatches 415; Indels 133; Gaps 24;

Qy 75 LPPYDFTALIRRGRTICSTVPAKAEKQSLFTECITTYNSDMLVNYKEDYGE 134  
Db 1 MPEMEDISVIGRRRTVOSTVEDAKRAQSLFVKECITYSDMHVNVYKXEDPFGD 60  
Qy 135 FROLNKKVKKDKLVHYHYEDERVDKEDDAASIGSGGITKQGLYKGNNSAISVTM 194  
Db 61 FRLPCKSLRPEKLPNHVFEIDEDCEKDESSLSQSGVYKQGLKHAANVSTITVTM 120  
Qy 195 RSFKRPFHLIQLDGSYKFEFLKDLQ--KEPKSIFLGPLYGVSRNNKYRPAFELKM 252  
Db 121 KVFRRRYVYQLDPGSIYLSYKDEKNSKSCITYIDACIDV--VQPKRRHAFELKM 179  
Qy 253 QDKSSYLLAADSEVMEEMWITLTKLIQLNFEAMQEKRRNG--DSHEDD---QSTLBSGG 308  
Db 180 LDKYSHYLAETBQEMBEWLTLLKLIQINDSLVQEKKTIVRTAQDDBTSSQGAENIM 239  
Qy 309 SGLD-STYRELAKARRE--IKL-KSESRVCLFYLDPAKDLDPSSAEPKYSFEKFG 364  
Db 240 ASLERSMHPLEMKYGRETEQNLKLSRGDGRNLSFSDSEVYRDPISGIBPDIKPEEKN 299  
Qy 365 KRILVKNDSLNFNLCCVAENEBGPTTVPEFFVTLSLFDIKYRNKISADPFVLDLHFSV 424  
Db 300 KRLVNCCHDLFNLIGQGDNAKGRPTNVEFFNLALFDVKNCKISADPFVLDLHFSV 359  
Qy 425 ROMIATTSPALMN-----GSGPETQSALRGILHEAMQYPRQIGIFSVCPPHDTLVARI 479  
Db 360 REMLWGSSTQLASGSPKGSPE--SYIHGIA--ESQLRAYIQGGIFSVTPNPHPEILVARI 416  
Qy 480 EKVLAGSTTHCAEPYKMSDSSKVAQKVLKNAKQACQRLGQYRMPMAAARLTFDAAGN 539  
Db 417 EKVLAGSTTHCAEPYKMSDSSKVAQKVLKNAKQACQRLGQYRMPMAAARLTFDAAGN 476  
Qy 540 LDKNARFSAIYRQDSNKLSDNDMLKLADFPKPKMAKLPVILGNLDTINDVSDPNNY 599



TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: CLASP-4  
US-09-978-244A-29

Query Match 55.8%; Score 6105.5; DB 9; Length 2008;  
Best Local Similarity 57.8%; Pred. No. 0;  
Matches 1197; Conservative 327; Mismatches 415; Indels 133; Gaps 24;

75 LPPYDFOALIRGCRITCTYPAKAEBAQSLPTTECITKYNSSMHLVNTKYEGSGE 134  
1 MPEMEDIISVIGRORRTVQSTVPEBAEKRAQSLPKYECIKYTSIDMHVNTKYBFSFD 60  
135 PQLPKNVVKLKLPHVYEVDEEVDKDDAASLGSGKGTGKGLYKGNMSAISYTM 194  
61 PMLPKSLRPEKIPNHVEIDEDCEKDDSSLCQKQGVTKQGLHAAVNSITITVM 120  
195 RSPKRRFPHLIQLDGGSYKFEFLKDLQ--KEPKGSIFLGFLYGVSFRRNKVRFAPFLM 252  
121 KVFERRFYLTQLPDGSYILNSYKDEKNSKSGCIYLDACIDV-VQCFMRHAFELM 179  
253 QKSSVTLADSEVEHEWITLANKLQUNFPAAMEKRG-DSEHDE---QSKLESG 308  
180 LDKSHYLAETQEEMEWLITLKKLIQINTDSLVEKKIYETVETADDETSQGAENIM 239  
309 SGLD-SYLPFLAKSAREAR--IKL-KSBSRVKLFYLDPAOKLDFSAPEVYSPEKFG 364  
240 ALEHSMPELMKYGHETQNLKLSGDRONLFSFDSEYQRDPSGIEPDIKPEEKN 299  
365 KAILVKNDSLSPNLQCCVANEGBPTTNEPFEVTLSPDIKYNKISADPHYDLNHFVS 424  
300 KQFLVNCBTLTNNIGQIDNAGPRTVNEPFINALFDVNCKISADPHYDLNHFVS 359  
425 RQMIATSPALNM-----GSGPRTGALAGILHEAMQYPKGISEVTCPHEDIFVARI 479  
360 RMLWGSSTQLASDGS PKSSPE--SYHGIA-ESQRLYIOGIFVTPHBEIFLVARI 416  
480 EKYLGSIITHCAPYMKSSDSKVAKVJLAKNAKQACORLGOYRMPAMARLTFKASGN 539  
417 EYLGQNIHCAEPYIKNSDPVKTAKOKRTAKQVCSRLGOYRMPAMARLTFKDTQGS 476  
540 LDKNAFSAIYRQDSNKLSDNDMLKLLADFRKEKXAKLVILGNLITIDNVSDDPNY 599  
477 LLDGGRFPLYQDSSKLSBIDLKLSHYKKREK-TKQIIPQGLNITVECPVDSLNC 535  
600 VNSSYITPKQFE-TGCKPTITFEVEEFVPCIPKHTOPYIITNHLVYPRKYLKDSQSF 658  
536 ITSSYVPLRPFKNCON--ITVEVEEFVPEMTKYCYPTIKYKHLVYVFLQKYSQKTF 593  
659 AKARNIACIEFKDSEEDSOPKCIYGRPGGVPTFRSAFAVLHNNHREFYDEIKEL 718  
594 AKARNIACVEFRDSDSDASALKCIYGRPGSVFTTNAYAVASHNNOPEFDEIKEL 653  
719 PTLQHEKHLTLTFPHVSCDSSKSTKRDVETQVGSWLPPLKDGKSVTSEQHPVS 778  
654 PHLHQKHLLTFPHVSCDSEINTKGTQKQDYEFTVGFAMVPLKDGKIIIFEQGLPVS 713  
779 AULPSCYLGQELGMRHAYGPETIKATVDSGKPLKSTHUVSTVYTTODGLHNPQYQCT 838  
714 AULPPLYLNLDAESNRQCNVDIKMVDGAKPLPKFSHLESTIYTODLHVHFFHHCQLI 773  
839 BSGAOLAGNELVYKLSLHAMEGHVIAFLPTLNLQDFVLIRATOE-EVAVNVTRVIIH 897  
774 QGSGKEVPBELIKYIKLCHAMEIQVMIQPLPVILMQLRVLNMTHEDDVPINCTWVLAH 833  
898 VVAQCHESGLSHLSYVYKAYKAEPYVASEKTVHEELTKSMTTILKPSADFLTSNKL 957  
834 IYKCHIEGLDYSIRFIKSPREKSAPOAQLIHELATMIALIKQSAFLSINKL 893  
958 KTSWPFVDVLISMAOHLIENSVKLLRNQRPASHTHAIVETVNMMLPHITQKFRDNE 1017  
894 KTSWPFELIAKSMATYLLDENKIKLPRGORPEETVHVHLSILLAIIPHVTIRVAEID 953

1018 ASKNANSLAVFIKRCFTFMDRGFEVFKOINNYISCFADGDPKTLFEYKFEFLVVCNHH 1077  
954 ESRNNVSLASTKRCGLTMDRGFTFNLINDYISGFSKDPKVLAVKFEFLQTCINHH 1013  
1078 YIPLNLPMPEFGKRIQRYODLQDLYSLTDFCRNHFVLGLLREVGNALQDFEVRLLAI 1137  
1014 YIPLNLPMFAKPKLQRVQDSVLEYSLEDECKHFLVGLLRETSIALQDVEIRYTAI 1073  
1138 SVLKLLIKHSPDDRYASRSHQARIATYLPFLGILNNQIRINRQVSPFVNAGMTVK 1197  
1074 SVLKLLIKHADVTRYQHKNOQAKLAQLYLPFVGLLLEIRQLACRD----- 1120  
1198 DESIALPAMNPLVTPQKSTLDSNLKDL-----GALSIGASPYTSTPINSV 1247  
1121 -----TISCAAMPNNSRDEPFCGFTSPANRBSLSTDKDTATGSPONGGI 1167  
1248 RNADSRGSLISTDSGNSLPERNSEKNSLDKHOQSSYTGNSVVRCDKLDQSEIKSLMCF 1307  
1168 KREDRGSGLI-PEGATGFPDQGTGENT-----RQGST-RSBSVQYKFLDQYEIRSLMXY 1221  
1308 LYIISMSDDALFTYWNKASTSELMDFTISBVCLHQYVNGKRYIA----- 1354  
1222 LYIVMISEDTLTYWNKVSPEELINILILEVCLFHRVYKGRNIAVHDAMLSKHFI 1281  
1355 -----RTGMHARLQOLGSLDLSNLTFNHSYGHSDADVHQSLLEANIATRYCL 1402  
1282 DRKSQTMALNRSSVMAQRIQHLSSLESSFTLHNSSTTERDRIHQVLLBENTTYEVL 1341  
1403 TALDLSLFTLAFKNQLLADHGNPLMKVFDVYLCPLOKHQSEFALANVFALASLYK 1462  
1342 TVLDITISFTQCFKQQLNNQGNHPLMKVFDHILAFKNGQSEVSLGHVFAFLAPISK 1401  
1463 PPSFTYEGRADMCALCYEILKCCNSKLSIRTEASQLLYPLMRNPDTYTGKSVRTHL 1522  
1402 PPSAFKRVNMCAPCYEVLKCCNSKLSIRTEASALLYPLMRNPFYTKRKTFLRTHL 1461  
1523 QYIISQULIADVIGSTRFQOQSLIINNANSRDLIKHTSPSSDVLDTKRITVMA 1582  
1462 QIITVSQLIADVLSGSRFQESLFTINNANSRPMKATAPRAVEVDLTRITVMA 1521  
1583 TAQMEKENDEPMLVDLOYSLSAKSYASTPELRTKWLDSMARIVHNGDLSBAAMCYVHT 1642  
1522 TAQMEHEKDEPMLIDLQYLSAKSYASTPELRTKWLDSMAKIHVNGDPSBAAMCYVHA 1581  
1643 ALVAEYLTRKCAVQMEPPLHSHSACLRBSRGVFRGCTAPFRVYTNIDBEAAMEDV 1702  
1582 ALVAEFLRRKK-----LFRNGCSAFKKITPNIIDEBGAMKEDA 1618  
1703 GMQDVHFNEDVLMELLQCADGLMKABERYBLIADYKLIIPYKGRDPBERLAHYDTLH 1762  
1619 GMQDVHYSBVLBELBECQDGLMKABERYELISELSKLIIVYKGRDFEKLTOYRRLH 1678  
1763 RAYSKYTEVMSGRRLLGTFRVRAFFGQAQYQFTSETDVEGFEZDEBGEKIYKBPUL 1822  
1679 GAYTILLEVMTKRRLLGTFRVRAFFGQS-----FFEBBDGKEYIYKBPUL 1724  
1823 TPLSISQRLKLYSDKRGSENVKMIQDSGVNPKDLSKATAYIOVTHVIPFDEKELQ 1882  
1725 TGLSEISRLVLYLQYKFKGTENVKLIQDSDYNAKELDPKTAHIQVTVKPYFDDKELTE 1784  
1883 RTKEPERSHNIIRPFEMFPTOTGRQGVGEQCKRRTILAIHCFPVYKRIIPMYQIH 1942  
1785 RTKEPERSHNIIRPFEMFPTOTGRQGVGEQCKRRTILAIHCFPVYKRIIPMYQIH 1844  
1943 TDLNPIEVALIDEMSKKVALBOLCGSAEVDMIKQLKQGSVSVVNAGPLAYARAFLD 2002  
1845 INLKRPIDATDEIKQTKAELQKLCSTVDVMIQQLKQGVNSVQVNAGPLAYARAFIND 1904  
2003 TINTKRYPNNKTKLKEVROVEACGALANBERLIKEDQLEIYQEMQANREMAKELSE 2062  
1905 SQASKYPPKQVSELKDMFRKFIQACSIALBENRLIKEDQVEYHGLKSNFRDVKELSD 1964

Mon Jul 14 19:53:42 2003

us-09-815-379-8.rapb

Page 20

```
QY      2063 IMHEQICPEEKTSLP--NSLHFNIAISGT 2091
        |::|||::|::|::|::|::|::|::|::|
DB      1965 ITHQI--LQEDTMSPMSTLTHTFCAISGT 1994
```

Search completed: July 14, 2003, 18:27:11  
Job time : 59.3964 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 ; Search time 46.6135 Seconds  
(without alignments)  
4345.419 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936  
Sequence: 1 MSQPLPLPASAETRRKFRAL.....ISGTPSTWVHGMTSSSVV 2107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR\_73: \*  
2: p1r1: \*  
3: p1r2: \*  
4: p1r3: \*  
5: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2888	26.4	738	2 T60486	gene tsg protein -
2	1977.5	18.1	2374	2 T31052	hypothetical prote
3	1664.5	15.2	2018	2 T34274	hypothetical prote
4	540	4.9	1105	2 A71430	hypothetical prote
5	429.5	3.9	1759	2 T18868	myoblast city prot
6	382	3.5	1970	2 T03284	myoblast city prot
7	279.5	2.6	2550	2 B53435	vesicular transport
8	232.5	2.1	1932	2 S53409	probable membrane
9	223.5	2.1	2819	2 A90551	conserved hypochet
10	224.5	2.1	831	2 T08611	hypothetical prote
11	222	2.0	2166	2 G70163	hypothetical prote
12	214	2.0	2401	2 T28676	rhodopsin protein -
13	213.5	2.0	3685	1 A27605	dyseroplin, muscle
14	212.5	1.9	2854	2 T14156	kinasein-related pr
15	204.5	1.9	843	2 T01438	hypothetical prote
16	203	1.9	2077	1 W2B224	240K tegument prot
17	198.5	1.8	1620	2 S61535	nucleotide-binding
18	197	1.8	2028	2 T08025	DNA-directed RNA p
19	195	1.8	2078	2 T09326	tegument protein -
20	195	1.8	3599	2 P90608	ABC transporter pe
21	194	1.8	3119	2 T18414	protein g377 - mal
22	193	1.8	1325	2 T42722	male-enhanced anti
23	192	1.8	3225	2 I52300	giantin - human
24	191	1.7	3259	1 A56539	giantin - human
25	189.5	1.7	2748	2 S57976	nuclear migration
26	189	1.7	2269	2 T28677	rhodopsin protein -
27	189	1.7	2297	2 AB2494	hypothetical prote
28	188.5	1.7	1939	2 T18372	repeat organellar
29	187.5	1.7	1778	2 JT0382	apolipoprotein B -

30	187.5	1.7	2629	2 I46569	apolipoprotein B -
31	187.5	1.7	2829	2 A42771	reticulocyte-bind
32	186	1.7	1818	1 S73852	hypothetical prote
33	186	1.7	2663	1 S28261	centromere protein
34	185	1.7	1557	2 T18412	lipid-binding prot
35	185	1.7	4464	2 D87755	protein T1812.4
36	184	1.7	840	2 T01357	hypothetical prote
37	184	1.7	1289	2 S67200	hypothetical prote
38	184	1.7	4385	2 T29042	hypothetical prote
39	183.5	1.7	1957	2 T38077	hypothetical colle
40	181	1.7	1837	2 T41023	probable nuclear p
41	181	1.7	2014	2 S46622	probable membrane
42	181	1.7	2649	2 A40937	bulbos pemphigoid
43	179.5	1.6	1963	1 MWKM	myosin heavy chain
44	179.5	1.6	1997	2 F71607	DNA helicase II BR
45	178.5	1.6	1211	2 S68251	phospholipase C, 1

## ALIGNMENTS

RESULT 1  
160486  
gene tsg protein - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999  
C:Accession: I60486  
R:Pinane, L.; Porcellini, A.; Avvedimento, V.E.; D'Esposito, F.; Felicitello, A.; Montice  
Mol. Biol. 13, 75-83, 1994  
A:Title: A novel thyroid transcript negatively regulated by tsh.  
A:Reference number: I60486  
A:Accession: I60486  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-738 <RES>  
A:Cross-references: EMBL:X68101; NID:G550419; PIDN:CAA48220.1; PID:G550420  
A:Genetics:  
A:Gene: tsg

Query Match	26.4%	Score 2888	DB 2	Length 738
Best Local Similarity	87.1%	Pred. No. 9.8e-159		
Matches 573	Conservative 23	Mismatches 36	Indels 26	Gaps 7
QY	1424	GNPLMKVDFVYLCFLQKHQSEETLKNVFTALRSILYKFPSTFEGRADMCALCYELL	1483	
DB	5	GHSLMKVDFVYLCFLQKHQSEETLKNVFTALRSILYKFPSTFEGRADMCALCYELL	64	
QY	1484	KCCNSKLSIRTEASQLLYFLMRNFDYTGKSPVRYTHLOVYISVSQIADVGIIGTRF	1543	
DB	65	KCCNSKLSIRTEASQLLYFLMRNFDYTGKSPVRYTHLOVYISVSQIADVGIIGTRF	124	
QY	1544	QOOSLIINNCANSDRLIIGTSPSSVVDLTGRIKRVLMATQMKHEHNDPEMLVDLOYSL	1603	
DB	125	QOOSLIINNCANSDRLIIGTSPSSVVDLTGRIKRVLMATQMKHEHNDPEMLVDLOYSL	184	
QY	1604	AKSVASTPELRTKWLDSMARIVHKNGDLSBAAMCVHTALVAAYLTTRKE--AVQMBEP	1660	
DB	185	AKSVASTPELRTKWLDSMARIVHKNGDLSBAAMCVHTALVAAYLTTRKEADLALQREBP	244	
QY	1661	LLPHSHACLRSSRGVFRQCTAFRVITPINIDEASNMEDVGMQDVAFNEDVLMELLEQ	1720	
DB	245	VFPVSHTSQCRKSRGCMFRQCTAFRVITPINIDEASNMEDVGMQDVAFNEDVLMELLEQ	304	
QY	1721	CADGLMKAEKRYE--LIADYKLIPIYKRDPEFLAHLYTTLHRAYSKVTFWHSGRL	1778	
DB	305	CADGLMKAEKRYE--LIADYKLIPIYKRDPEFLAHLYTTLHRAYSKVTFWHSGRL	359	
QY	1779	LGTYPRV--AFPGQAQYQFTDSEVDGCFEDEDGKEYIKKPLQPLSEISQRLKLY	1836	
DB	360	AGSNLPLGGLPGQ-----GFPEDEDGKEYIKKPLQPLSEISQRLKLY	405	
QY	1837	SDKFSSENVKMIQDSGVNPKDLSKAYIVQTHVIPFDEKELQERKTEFERSHNRFP	1896	

Db 406 SDKSGSENVKMIQDSKVNPKDLSKFAVIQVTHVTFPEDEKELQERKTEFERCHNIRRF 465  
Qy 1897 MEMEPTQGTGKGGVEEOCKRRTITLAIHCFFPYVKRIIPVMYQHHTDNPTEVAIDENS 1956  
Db 466 MEMEPTQGTGKGGVEEOCKRRTITLAIHCFFPYVKRIIPVMYQHHTDNPTEVAIDENS 525  
Qy 1957 KKVAEIROLCSSAEVDMIKQLKLGSSVSQVNAAGLAAVAAFLDDTNTKRYPDNKVKLL 2016  
Db 526 KKVAEIROLCSSAEVDMIKQLKLGSSVSQVNAAGLAAVAAFLDDTNTKRYPDNKVKLL 585  
Qy 2017 KEVFRQFVACCGALVNNRLIKEDOLEVOEEMKANYREMAKELSIIMEQICPLEEK 2074  
Db 586 KEVFRQFVACCGALVNNRLIKEDOLEVOEEMKANYREMAKELSIIMEQICPLEEK 643

## RESULT 2

T21052  
hypothetical protein F22G12.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 29-Oct-1999  
C:Accession: T21052; T21281  
R:Kerhaw, J.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19365  
A:Accession: T21052  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2374 <M11>  
A:Cross-references: EMBL:Z81066; PIDN:CAM02974.1; GSPDB:GN00019; CESP:F22G12.5  
A:Experimental source: clone F17B5  
R:Liernard, N.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19400  
A:Accession: T21281  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2374 <M12>  
A:Cross-references: EMBL:Z92831; PIDN:CAM07369.1; GSPDB:GN00019; CESP:F22G12.5  
A:Experimental source: clone F22G12  
A:Gene: CESP:F22G12.5  
A:Map position: 1  
A:Intome: 47/2; 75/3; 114/3; 172/2; 230/3; 322/1; 376/2; 439/1; 653/2; 731/3; 758/1; 80  
/2; 2139/1; 2258/3; 2298/2; 2344/3

Query Match 18.1%; Score 1977.5; DB 2; Length 2374;  
Best Local Similarity 25.6%; Pred. No. 2.4e-105;  
Matches 628; Conservative 416; Mismatches 918; Indels 489; Gaps 73;

Qy 9 ASAEKTRKFTALSKPCTAAELROSVEVGVSLAKPLIEPLDYENIVQKTCQIIND 68  
Db 41 ATSQARKFTAGVGRPELAEARDAVAAALVAQNPKEPISSEVNNKEFINERSIQLEND 100  
Qy 69 CUREMILFFYDFQTAIILROGRYICSTYPAKABEASQSLPTECIKTYSNMHLVNYK 128  
Db 101 POREIYLFPRDIEESQOTCEVAPVYKPDINDIKAKMLTREARLFYTTAHSISIVNY 160  
Qy 129 EDYSGEF-RQLPKVVKLKLPHVYEVDEVDKDDASISQSGKITKGMVYKNN 187  
Db 161 SKPSGDYSANNTSSAIENETLSLVEESMALE-DERAAFGASSNGAV---GVVXEGNLN 216  
Qy 188 SAIS-----VTMSKFRKRFPHLITQDGSYKFEFLDQKEPKGSIPLGELYGVSPR 239  
Db 217 VLKSDPPTLFDMLKSSSARRYCILARHGSECTFIRKTAQEAR-MLPLKVASAOK 274  
Qy 240 NKKVRFAELKQO--DKSYLLADSEVMEEMITLL----- 275  
Db 275 STRKGTVEIRSDADKAAIILLESBETQMLEMVAALTAALANKEDAISCELDYK 334  
Qy 276 ---NKILQINFEAMQEKNGSDHEDEOSKLEGGSGGLDYLPELAKSAREABEIKLKS- 331  
Db 335 GIGSKLDDVNAATSSSTVGGGRSKGADTESIGSEDSSNCW-----NSANFEPKKLKNC 389

Qy 332 ---ESRVKL-FYLDPAQKLDSSS---AEPVKSFEKFGK-RIL----- 368  
Db 390 TFFFRKRLDIFYSKTRKSPERSENLFKELEKIQGYIEKFGKTLIPFTRTSEMQWRG 449  
Qy 369 -----V 369  
Db 450 RNSAARALOPPIVDRNNFSLYRLQPLPQSPHESITLPGNSSKRGADAPIFSSMSPV 509  
Qy 370 KCONDSPNLOCCVAB-----NBEGPTVNEPFFVLSLFD 404  
Db 510 KKRKTSGLKRGKGAOSSLSSTGTPLSLDPFHGLHLPLNLSGLVLOQIEFFVAFPFVD 569  
Qy 405 IKYNNKISADPFAVDNHSVQMIAATSPALMNGSPETOSALRGILHEAMQYKQIGF 464  
Db 570 ALFARRACBEFQILVN---GDELHLPVF--DSCKYNIGIRSLVDRASRALQLP 623  
Qy 465 SVTCHPDIPLVAREKVLQSGITTCAPFPYKSSSVKVAQVYLKNAQACORLQGYRMP 524  
Db 624 PTTVNRDLMTVCYCRDRL--SADTAELYNKSSDPRAVAKLOQTVSMSLTRLGHRQR 681  
Qy 525 FAMAARTLPKASGNLDKNA---RPSAIYRQDSNKLSDNMLKLADPRKPEKNAKLPVI 581  
Db 682 FAWTKAPLPPELRKSTHQRSAARRLCFLHQLQHTLDQOKYOML--FNCTDRQSKMILP 739  
Qy 582 LGNLDITIDNVS--DFPNVYNSYIPTRQF--ETCSKTPITFEVEEFVPCI PKHTQRYT 637  
Db 740 NATVGSVNTVTRASVDYLNRVNPSLXPJNPMKPEDPMTPIPVFQSGFG---DQSGQPH 796  
Qy 638 IYTNHLVYYPKTKDKDSQKSPAKANLACIEFQDSDEDDQPLKICIGR--PGFPVETR 695  
Db 797 SLTNLLYIYPLALKDSQKAFSKANISCTYRFRNG--EEAIPKAMVDRMSAQP-YCI 853  
Qy 636 SAFAAVLHHQNPFEYDEIKIELPTQLEKHHLLTFPHVSCDSSKSTKRDVETQV 755  
Db 854 SSTCAVQHQQNPVGEEMKQPLPLNTSDBLHPFSHISVAGNS--NLKASESTETPI 911  
Qy 756 GYSWLPPL--KGRVVTSEQH--IPVSAVLPQGYLQYQELQMGF--HYGEPIKVVDSGK 808  
Db 912 GYSWLPVWKCDRLVMEDEQFALPVALIDPNNYYSKPTGMGKDETLTSEVRVWD--QK 970  
Qy 809 PLKSTHLSVTVYQDQHLNFPQYCKQTSAGNALGNELVKYLSLHAEHGMIMFL 868  
Db 971 PLFRRRLVSSVFTDQKLTFFQACRL--SAGGIIIGDAADSKLSLSDVPFRLLVYL 1028  
Qy 869 FTILNQLFRVLTRAQOEVAANVTREIIVHVAQCEBGLJESHSYVKAAY----AEPY 924  
Db 1029 PIVGLRLAAILPQATBGLATATLRSISIDCMCCQNKQPVVRRFVSHSDSARQF 1088  
Qy 925 VASEYKTVHEBLTYSMTILKPSADFLTSN-----KLKYSWFFPDVLKSMQHLIENS 979  
Db 1089 VSHL-TTIYSAILCRHLPTLMRE---LQSEBQALQVLYQLVLTDAVTKSMAQTMCSQ 1143  
Qy 980 KKKLARNQRPASTHNAVETVYNNMLPHITQKFRNPASAKNHSLSLVFIKCTFMDR 1039  
Db 1144 LNKTSQRDRFOEILQGVQVLEGVPOIVLKHMEKREBSCANLALAYFTRFAMSFVDR 1203  
Qy 1040 GFVFKQINNYISCFAPGDKPLTFEYKFEFLVVCNHEHYIPLNTPM----- 1085  
Db 1204 GVFWFHIPIYSRLDDTTRALRYKTDLLBILCHENHVNLPVLLNMAAQIQRLYYS 1263  
Qy 1086 -----PFGKG-----RIQ 1093  
Db 1264 GGVVDQTQMTNGTGLSGGYCSLSRSRSTVVNNIRASGSGFLRFNFQNTPTLETNETD 1323  
Qy 1094 RYQDLQDYSLTDECRNHPVLGILLRVRGALQEFREVRLLAISVLYKLLIKSPDRY 1153  
Db 1324 RYASGSGEMHLSPAQAQNHFTVGLMOELVACIBETQYRKRPISLNLNLLAKSHFDRY 1383  
Qy 1154 ASRSHQAIATLYPLFGLILEYVQRIIVRVVS-----PEPVNAGMTVKQESLAL 1203  
Db 1384 GDMTQRIIMLYAPVLTALPDHLHERFSDSIDVDVATPTGYRSRPTSGGLP----- 1436  
Qy 1204 PAVNPLVTPQKSGTLDSNLKDLGLAISGLASPYTTSTPININSVNNADSGSLISTD--- 1260

```

Db      1437 PSSSTTKTISKRTVNR--YTDVQPSRG--SPIRTSSP--VPSQTTPSRPPQMPPPPA 1491
Qy      1261 SGNLSLPERNSEKSNLSDKHQSSSTLGNVRCQDQSEIKSLMCEFLYILMSMDALF 1320
Db      1492 SGTITPPPPS-----SSTPLVEKLTEDDEIDLLICVITLQMPKRIIA 1535
Qy      1321 TYW--NKASTSE-----LMDPF--TISEVCHQF-----QYMGKVIAR 1355
Db      1536 ALMTENDGANNAEMKIRLELIYDVPLFEMALQNGVEYVGKIMLVKKRRGKGNHAR 1595
Qy      1356 TGMTHARLQOL-----GSLDNLST--FNHSGHSDAD-----VLHOSLIE 1393
Db      1596 RTAANSKTSLEFTLNLPSRISSGSTPNRISAEWMAAGABDDPNSSGTTNPDPFVLOL 1655
Qy      1394 ANIATEVCITALDTLSLFLAPNOLAD--HGHNPD-----LMKKVDVYLCELOKQSQST 1447
Db      1656 VNLQSEVALIVDVAO---TPAHQLASSQRHHCQSDALFHSILALHLRLADEHWSST 1711
Qy      1448 ALKNVFTALRSLLYKPPSTFYE--GRADMCAALCYEILKCNKSLASIRTEASQLYFLMR 1506
Db      1712 VRLHYIAGALFVNLFARLPEEGPLEPLYMILEKVLQMASRLPAVQAAAAALLQLILR 1771
Qy      1507 NNPD---YTGKKSFTVTHLQVITISQLADVIGIGTR-----PQO 1545
Db      1772 NGYEVAOGYFAQOVLAVSVSPSSKINQOAAATRKGVSSERLGRPGSQGVALARLLGFG- 1830
Qy      1546 SLSTIINNCANSRLI-----GHTSPSDVNDLTTRIRTVLMATQMKEMHEDREM 1595
Db      1831 --SVLSNSAPFERGLAAVEALVDTRKATSFDLALDLRLQGVMTATVALDOAADPFR 1888
Qy      1596 LVDLOYSLAKSYASTBELRKTMLDSMARIVKNGDISEAMCYVHTALVAEYL--TRKE- 1653
Db      1889 LADLHQLADSYRGSAAALSAMFDTLAEIYEDRWAEASVCHASVALIAABELERKEL 1948
Qy      1654 AVQMEPRLPHSHSACLRSRGVFRQGTAFRVITPNI DE--EASMEEDVG--MDVHENE 1711
Db      1949 EVDWR-----VDFWIMNRJAETEOSGCGAGASQVAPAGFT 1983
Qy      1712 DVLMLLEOCADGLMKEERYELLADIYKLIPIYERKRPRLAHLYDLTHAAYSVTEV 1771
Db      1984 DNLGAKIDTAAALMAERFEAVGPLYRILIVPLEKMTFTSLVSVAELQOYTSRAAEV 2043
Qy      1772 MHSGRLLGTFRVAVFGQAQOYFTDSETVGFEDEGKEYIYKEPKLTPLSISQR 1831
Db      2044 RSSGRHLGAYFRVRNG-----ERNHGSHTNDWIYREGSLISLAFALF 2089
Qy      1832 LKLKYSDFKGSBNVKMIQDSKVNPRDLDSKYAYIQVTAVIPFEDEKE--LQERKTEPER 1889
Db      2090 IREKQORQVGHDRVQ--IEANEOILSKIDPTVAVYQITFVEPSIPAAAGIADQHRNDFLV 2148
Qy      1890 SHNIRFEMEMFTQVGK--RQSGVEEQCKRTIL-----TAHCFPPYKRI 1935
Db      2149 HTNLSEFSECATTEBERKVSKEPAHEQCLKRTYLRVSPSVSEDSRAATGFPATRRLL 2208
Qy      1936 PYMGOHTDNLPIEVALIDEMSKKVALRQLCSA---EVDMIKLOLKQGSVSVQVNG 1991
Db      2209 PVIASHPEQSPLEFPAQCLTKABOIRKTIANAASGRQLDVKQDOLLQGAVALPRTVNG 2268
Qy      1992 PLAYAVLFDITNTKRYPDNKYKLLKEVFRQVEACGALAVNERLLIKEDOLEYOEMKA 2051
Db      2269 PLAYAEVFTKEQREYGDGLVKLRSEFRNLNNSCOLAIEANASAIQSDQTYHEVLVS 2328
Qy      2052 NYREMAKELSEIMHEICPL-----EKTSTVL--NSLHFNASG 2090
Db      2329 SF-----DANHERLQTFPGASLRGSLPEQSSSQFPPRSAMHILDMWG 2371

```

RESULT 3  
 T34274  
 hypothetical protein F46H5.4 - Caenorhabditis elegans  
 C1:Species: Caenorhabditis elegans  
 C1:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

```

C1:Accession: T34274
R1:Nhan, M.
submitted to the EMBL Data Library, November 1995
A1:Description: The sequence of C. elegans coemid F46H5.
A1:Reference number: 221498
A1:Accession: T34274
A1:Accession: preliminary; translated from GB/EMBL/DBJ
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: DNA
A1:Residues: 1-2018 <NHA>
A1:Cross-references: EMBL:U41543; PIDN:AA837023.1; GSPDB:GN00028; CESP:F46H5.4
A1:Experimental source: strain Bristol N2; clone F46H5
C1:Gene: CESP:F46H5.4
A1:Map position: X
A1:Intron: 16/2; 52/3; 87/2; 116/2; 138/2; 203/1; 265/3; 317/2; 337/3; 378/1; 428/1; 482/3; 1491/3; 1560/2; 1632/2; 1753/3; 1830/2; 1866/2; 1927/3

Query Match      15.2% Score 1664.5; DB 2; Length 2018;
Best Local Similarity 23.1%; Pred. No. 2.2e-87;
Matches 546; Conservative 406; Mismatches 746; Indels 665; Gaps 73;

8 PASAETRTKTRALSKRTPAELROSV-----SEVVG-----SVLAKPRL 48
Db      3 PADGK-RAPVKKRSIV-TASVVRKHVIGSLPHIRLSGDNGLDAVSMEXIQFTPE-- 58
Qy      49 IEPLDENVIVQKTKQILNDCLREML-----LFPY--DDFOATILRQO--RYICST 96
Db      59 VPPVVEALLDQRT-----CLYSSPKSKSLPEFVSDVAVETTTROGLTDHNTT 113
Qy      97 VPAAKEEBAQSLFTECIKTYNSDMHLYNKYEDYSGEFROLPNKVAKLDPVHYEVD 156
Db      114 IP-NIENH-----VADIGFCYCDNPSLVNRKY----- 139
Qy      157 EVDVDEDAASISGQKGIITGCMYXGNMNASISVTMRSPRRRPFHLIQLGDSYKPEF 216
Db      140 -----AQGTED-----IKRMLEKLTALRSRLPQIFH-----NG 170
Qy      217 LKDLQEKPKGSIFLFLGVSEFRNKKVRFAFE-----LKMQ--LKSSYLLAADSEVM 268
Db      171 LRIMREP-STISID---GGSFDTIDPTKCAEQGYVMLKSAIDKDTLYSMLSKRV 226
Qy      269 EEWITLNLKILQINFEAAMQEKRNDSHEDBQSLBEGSGGLDYSILPELAKSAREATIK 328
Db      227 KIFMNCIN-----EEEKSWKMERG-----VPRLP----- 251
Qy      329 LKSESRVLFYIDPDAQKLDPSAPPEVKSPEBKGRKILVKCNDLSRLLQCVANENRG 388
Db      252 -EQEETPKLFLV-----FVEKAAADP----- 270
Qy      369 PTTNVEPFFVTLSPFDIKYKNTISADPFVNDLHFSVRQMIATTSPALNNGSGPTQSALR 448
Db      271 ---FPEPLFASALYDIDKORQKVTSMTFNADHKLDMLSHQHPFNNY----- 318
Qy      449 GILHEAQAQYKQGISYTCPPHPIFLVARIKVL--QGSITHCABPYNKSDDSKVAQKV 507
Db      319 -----MQVLFNVTGKLEDMFLVKLEKVLQGDVFNSESPYGTQDENNM--EKL 366
Qy      508 LKNAQAQCORIGQYMPRAW----- 527
Db      367 BRPAERKNCORIGAYSPGFOVIDLQRIYKANVSTGASSPARRTPMMMSQCTTASGAVL 426
Qy      528 -----AARTLFPDAEAGNDK----- 542
Db      427 TTAGSQSDQDQGSITSADRTSIASMSGSLTRRGSGSALATVSRVATPLTKRFAPAVSNLP 486
Qy      543 -----NARFSAIYRQDSNKLNDMDLKLADFRKTRKRAKLPVILGNLDI 587
Db      487 TSOEVPENIEMWPSCNLFSSFRIRQEGDGTDEDIYRICEKRRKNGVKHKKGFNFLEL 546
Qy      588 TI--DNVSDPFPNPNYSSYIPKQPEKTSKPIPFVEVEFPFCIKKHQPTI--ITNNLY 644
Db      547 TLGANSKKEVQSHGSLN-----TLNSERVHHEAME--IPVYASLNSKYQNVIF 594

```



```

Db      127 PEMOLKESVRCIMQIFDCLLTVHERCKKGLSLARLNSLAFF---CYDUL----- 176
Qy      1045 QNNNTYSCAPGDPKTLFEYKFEPLAVGNHNYIPLNPMFPGKRIQRDYDLDYSL 1104
Db      177 ---YI---EPCOVE----- 186
Qy      1105 TDFECNHLVGLLREVEGTAQEPREVLIAISVKNLLIKHSPDRYASHORAIAT 1164
Db      187 -----LGRILVILLCHHEPARQAKGKUYINO 216
Qy      1165 IYLPFLGLIENVQIRINVDVSPPVNAGMTYKDESL--ALPAVNDLVPQKSTIDNSL 1222
Db      217 IYFPFVGQIL-----DEMPVFYNLATEKREVLIGVLOIVRLDSTSLVKMAQOSI 267
Qy      1223 -----HKDLGASISGASPYTTSTPNI NSVRNADSGSLISTDSGNS 1264
Db      268 ATRLTYFKMEBCLLIFHHKKAADSLG-----GNNSKGPV---SEGAG 308
Qy      1265 IPERNSKSNSLDKHQSSSTLANSVVRCDLQDSIKSLIMGFLYILKS--MSDGLFTYW 1323
Db      309 SPKYSERLSPAINNY--LSEASRQEVRLSEV-----TLISFGGLTKTKLECPITLFAVY 360
Qy      1324 -----NKASTSELM-----DFTTSEVCLHQFOYMGKRYIARTGMMHARLQOUGSLDN 1371
Db      361 MNLGNAQWPSBSLLQLLEGTPNGYLMGRVNSQLASPSQPSYSLREALAQAGSSRIGASAO 420
Qy      1372 SLTFNHSYGHSDADVHQ--SLEENIATEVCLTALD--TLSLFTLAFKQQLADHG---- 1424
Db      421 ALR-----ESLHPILIRQKLEMBEENAVATSLQVLEITENSSMAASHNITDYGKLD 474
Qy      1425 -----HNPLMKVDPVYLCLFOKHQSELTAKNV--FTALRSILYKF 1463
Db      475 ITTILTSFESRNOSLAFMKAFPIPIRIRIDLGATLMAENBRFLKQIFHLIRLAVYN 534
Qy      1464 PSTYEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLKRNFDYTGKKSFWYTHLQ 1523
Db      535 DSVKRAVIGLQILVVKHLESCKOSL-----FLT-----IRLAL 570
Qy      1524 VIISQSLADY-----VOIGSTRFOQSSIIINNCANSRLIKHTSPSD--- 1568
Db      571 LTTITSELSMDVQVTHMKSNDTLBESGEARRLOQSSENADEKSNVLLRECGLPDITLL 630
Qy      1569 -----VKDLTKRIRTYLATAQKHEHNPMLVLDQ-----YSLAKSY 1607
Db      631 IIPKFTENRMSMAEYKHLSDSL--VLADASLGHLLSLSVAMBRYYAAESFYKLGMAF 688
Qy      1608 ASTPELRKTLWDSMARIVHKNGLSEAMCYVHTALVAEYLTTRKAAVQWEPPLPHSHS 1667
Db      689 AEPVPLHIMWLLHLCDHQEMQSWAAQCAVAVAGVIMQALVARNDGVMS-----KDHV 743
Qy      1668 ACLRSRGVFRQGCAPRAVITPNDIEBSMMEDVGMQOVHNEVDLMLLQCADGLM- 1726
Db      744 SALRKI-----CP--MVSGETTEASABEVEYGASKULTVSAVYTL--QLANKFS 791
Qy      1727 KAERYELADIYKLIPIYKERRDFERLAHLVDTLHRAYSKYTEVHSGRLL--GTYPVY 1785
Db      792 QMBELHFCASILEYIPIVYKSKAYQKAKCHTLNLNYESIIDOSNPIPTIDATYIV 851
Qy      1786 AFGGAQAQYQTDSETDVEGFEDDEGKEYIKBPULTPLSISORLLKLYSDKFGSENV 1845
Db      852 GFYGEK-----FGKLDKREYVYREPRDVLGDIIMELSHIYESRMDSNHI 896
Qy      1846 -KMIQDSGVNPKDLDISKAYIQTHTVIFPFDEKLEQEKTEPERSHNIRRMFEWPFQ 1904
Db      897 LHIIPDSROVKAEDDQAGVCYLOITAVDVMEDEDGSR--ERISFL-----S 943
Qy      1905 TGRQGVVEEGCRATILTAICFPYVKRIPVYQHTDNLPIVAIDEMSKVALELQ 1964
Db      944 TGSVARTGS-----FPALVNRLLVTTSESJEFSPVENAIGMIETRTALRN 991
Qy      1965 LCS---SAEVDMI---KLQLKQSGSVSVQVNAGPLAVARAFLDNTTRYPDNKVKLK 2017

```

```

Db      992 ELEBRSSDGDHLPRLQSLQRILOGSVAOVNCGVLSVCTAFLSGEPATRLRSOELQLI 1051
Qy      2018 EYVROFEACGQALAVNERLIKEDQLEYOEEMKANYREMAKXLS 2061
Db      1052 AALLEFMAVCKRAIVHFRLLIGEBDEPHYTQLVNGFQSLTABELS 1095

RESULT 5
T18868
Annoblast city protein homolog - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C.Accession: T18868; T20461
R.Cummings, P.
submitted to the EMBL Data Library, October 1996
A.Reference number: Z19034
A.Accession: T18868
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1759 <W12>
A.Cross-references: EMBL:Z81032; PIDN:CAB02723.1; GSPDB:GN00022; CESP:C02F4.1
A.Experimental source: clone C02F4
R.Wild, A.
submitted to the EMBL Data Library, October 1996
A.Reference number: Z19278
A.Accession: T20461
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1759 <W12>
A.Cross-references: EMBL:Z81054; PIDN:CAB02887.1; GSPDB:GN00022; CESP:C02F4.1
A.Experimental source: clone F01D4
C.Genetics:
A.Gene: CESP:C02F4.1
A.Map position: 4
A.Introns: 11/1; 130/1; 194/3; 240/2; 313/1; 676/1; 702/2; 746/2; 841/3; 895/3; 1001/2;

Query Match      3.9%; Score 429.5; DB 2; Length 1759;
Best Local Similarity 18.8%; Pred. No. 3.5e-16;
Matches 389; Conservative 328; Mismatches 758; Indels 589; Gaps 92;

Qy      201 FPHLIQLDGSYKFFFLKDLQKPKGSIFLG-----FLYGVSP--RNKVRRA 247
Db      7 YYAIVQ-----YTFEDPLPSCDAPRLLEFIDRLCVYQGHGHWGGRKFKDEKNGCGIFP 61
Qy      248 FE-LKMDKSSYTLADSEVMEEMITILN-----KIQLNFEAAQKXNGDSHD----- 298
Db      62 LAYVOIVKSMFVSTSDGLVVDLISRVINEMWTIKEL-----WVETTRIGSFEDLMD 116
Qy      299 -DE-----OSKLESGSGGLD-----SYPELAKSARAERIKLSSESRVLFYLDPA 344
Db      117 FNEILLIKTKIESGCIPIEBLSKLRARVSKLVDRGNTLIGQDVIRNDEGVPL-----DV 171
Qy      345 OKLDPSASPEVYKSEFEKFGRIILVKNDSLFNLOCCVAENBEGFTTVBPFVTLSPD 404
Db      172 ESLSL-----LRTYEAHISK-----GRVQSLRREKRENTIN--DSFSLLSIKS 215
Qy      405 IKYNNKISADPHVDNHSVRQMIATTSPALM--NGSGPERQSLRGLHLEAAQYPRQG 462
Db      216 VELHCKYSCESISLYLDL-KMFTTDSYTPLMNSGSKHTDLNKAFLTPDAKEDIQK 274
Qy      463 IFSVTCPPHPIFLVARIKVLQGSIT-----HCAEPYKKSDDSKVAQKV 507
Db      275 YLMTV-----RVYHVSPISSNATMKHGHHEATIKTYCRGSY--ASDIMMSISLF 325
Qy      508 LKNKAAQACQRLGQVRMPAMARTLFKQASGLDNAR--FAAIYQDSNKLSDMDLKL 566
Db      326 LAGVGEAKS-----RVIFLNRBEPPLSLKAYHATYNI PKN--LSNDMEFTKL 372
Qy      567 ADFRKEQMAKLPVILGDLITIDNVSSDPNNYNSVYIIPKQPTCKSPITFEVEEV 626
Db      373 ISTQ-----LVPGN-----VSOIKARHPLP--SHCPAAILLRADRTAVSIDSR-- 415
Qy      627 PCIPHTQPTIYTNHLTVYVYKLYDSQKSAKARNAIACIEFQSD---EEDSQPLKC 683

```

```

Db 416 -----NMVITLMOAELSGSS-----DNIEARLHVSNNGHVENVEFIS 458
Qy 684 IYRPGGVP/PTRAFAVAHHHONPEFYDEIKETLQJLHEKHHLLTFPHSCDSSG 743
Db 459 VTGSQSLSTYK-----SIVYHTDKPMWTEPIKALSCHADYLLILYSSKAYKXP 514
Qy 744 STKRDUVETQVGYSWLPLKDG-----RVVTSQHI PVS-----ANLPGYLGQE-L 791
Db 515 EKGPFAIAHVQILRS-SALLCOSEHDLAVYKINDPSHPDESVMHMLPDRRTIKESI 573
Qy 792 GNGRHYGPEIKWVDGKPLKISTHLVSTVYTDQHLNFPQY---CQTESGAQLGNE 848
Db 574 GSAKPHSGOFSLSB--KSFWMISTHSCSSMLTQNEHLNVLNWRVNCVMLTSLVALAOP 631
Qy 849 LVKYLKSLHAMEGHVIAFLPTILNOLPVL-TRAQOEAVAVNTVIIHVAQCHE--- 904
Db 632 I-----GOTEHEMIRFPSHLLDALPEIMHORESEKV---VPDVIAVLRICEEPRH 680
Qy 905 -----EGLESHLR--SYVKYAYK-----AEYVASEYKTVHEBLTKSMITILKPSADFLTS 953
Db 681 YPQAAKIFEVYLLKRFSTSAALKILKQMLHVIYSDSEDSNEK-ARNAFVMGSLFVLVYV 739
Qy 954 NKLLKTSWFFPDVYLKSMQHLIE--NSKYKLRNORFPASVHH---AVETVNMMLPH 1007
Db 740 SKKCGIKFEFDDEFTYKQYLFEPKSLVALMSEKKAKMTVQNTALKSIPITIDLL 796
Qy 1008 ITQKFRDNPEASNAHSLAVFIKRCFTFMDRGFVQKQINNYISCAPDPTLFEYKPE 1067
Db 797 -----NESDVSSEPNLCGFIVDMNNGSNITVERLNFLA-----QIVTRF- 839
Qy 1068 FLRVVGNHEHYLP-LNLPMFPGKRIQRYODQLDLSLTDEFCRNHFLVGLLREVGYL 1126
Db 840 FSLSTICREQLLIPCLQIALDI-----IQIDMAT-----EKGEPA 874
Qy 1127 QERREVLIAISVLKVLIKHSPDRYASRSHOARIATLYPLGILLIENQORINRVDVS 1186
Db 875 DRAECASIIAILELFLDAKSGEENARELTSFLIMVYRPLVQAMIR----- 923
Qy 1187 PFPVNGMTVYKQESLALPAVNPLVTQKGSTLDNSLHKDLAIGSIASPYTSTPNINS 1246
Db 924 -----VIHDDKHTD----- 932
Qy 1247 VERNADRG-----SLISTDSGNSLPERNSEKNSLDKHQOOSTLGNSVVRCDKLDQSE 1299
Db 933 ---DARGHGFVITALLDKMSAQMFSEYEESSSIDIKRDLMEVWQMI--RDLNRA 987
Qy 1300 IKSLLMCFYILKMSDDALLFTYWKASTSELMDFITISEVCLHOF-QYMGK--RYIART 1356
Db 988 PFTWMDMIMLQNKVIHKSLL-----RFVMSAVOTFFSNKFCVEMWREYMTVVSFVTOE 1042
Qy 1357 GMMHARLQOLGSLDNLTFNHSYGHSDADVLHQSLLEANIATEVCITADLTSLFLARK 1416
Db 1043 G-----LNSKHEMKKEDMDRIQLRKA-----AAQDLASNMFLRLTP 1079
Qy 1417 NQLLADHGNPLMKKVPDYVLFCLOHQOSETLKNVFTALRSILYKFPSTFEGRADMQA 1476
Db 1080 SQKLL--NYIPSMIGSF-LKVSIVDDEDRERATPIFFDMQREY----- 1120
Qy 1477 ALCYELKCCNSKLSIRTE-ASQLLYFLMRNNFDYTKGKSFYRTHLOVITISVQLIADV 1535
Db 1121 -----NTSASRSFTEFASSELVSOLDTVNDQHSATKGFKEHFRQSLITLCOSEKEL 1170
Qy 1536 VGIIGTRFOOSLTIINNCANDR-----LTKHTSFSDVDVDTLTKRIRITVLMATAQMEHEN 1591
Db 1171 MANGSELLERI-----DRLLTALLIYHEVAS--KSLVECVDSLSRTVQLMRYYN 1219
Qy 1592 DEMLVDLOYSLAKSYASTPELRKTLDMEMARIHVNGDLSEAMCYV-HVTALVAEYLT 1650
Db 1220 -----QYSHKELVVK-----YIYKLYDLHISYGNKIEAKTLLNHAHMLTFD--- 1261
Qy 1651 KREAV-QMEBPLLIHSHSACLRRSGGVFRQGTAFRTITPNI DEEASMEVDVGMQDVF 1709

```

```

Db 1262 -DDALPQW---LIARS-----LNRHR-----EL 1280
Qy 1710 NEDVLMELLECQADLMKABREYELADIYKLIIPYK-RRDFEPLAHLDTLHRAYSKV 1768
Db 1281 HQQLKEDLMEBAGNLFSGEDMEDALLVYNQVLVYQNIIMDYDLKAGLLOKIAOYTSI 1340
Qy 1769 TEVMSHGRLLGTFRVAFFGQAAQOYFTDSETDVEGFEDEDEGKEYIKYKPLPLSEI 1828
Db 1341 SRT-----BRAVYVYLLVAFYQO-----GFPAYLNGHKFVFRSEKLEMHGEF 1382
Qy 1829 SQRLKLYSDKFGSENVKMIQDSKVNPKDLSKATYIYQVTHV-----FFDEKELOE 1882
Db 1383 MORIMKVY-----DNPKIMKTDP-PLHVDSPERYIQVFNIDIDIGCGSENNPEVKP 1435
Qy 1883 RKEFERSHNIRFRPFMEPFTQGRQGVQECRRTILTAHCFP-----YVKKR 1934
Db 1436 YKKTFRYYNIQTEYSK-----VEE--RKDTKMTSID--PSEEFMRMLVRR 1480
Qy 1935 IPVMYQHHITDL-----NPTEVADMSKVAELRQLCSSAE---VDMTKL 1976
Db 1481 IKTADSLPTDLRFEIIVELSDPIYVTPLONAVQKQKNGKELNETPAASNPFDKLL 1540
Qy 1977 QLKLOGSVVOVNAQPLAARAFLLDITNTKXYPDNKVKLLKEVRFQFVACGQALAVNR 2036
Db 1541 SRDILGVVSAVMGQ-----VKNY-----EVF--FTBACRNICEGB- 1575
Qy 2037 LIKEDQLEFYQEMKANYREMAKELSEIMHEOI-----CPLEEK--TSVLPS 2081
Db 1576 -----OSVIMELSLIIEVVEILEYCCYHABSCQGEARAINMPLPH 1618
Qy 2082 LHIFNAISGTPSTVWAGMTSSSS 2105
Db 1619 ANI--RSLSPDSDSINDGMWTKS 1640

```

```

RESULT 6
T03284
myoblast city protein - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 17-Nov-2000
C.Accession: T03284
R.Brickson, M.R.S.; Galletta, B.J.; Abmayr, S.M.
J. Cell Biol. 138, 589-603, 1997
A.Title: Drosophila myoblast city encodes a conserved protein that is essential for myob
A.Reference number: 214881; MUID:97392689; PMID:9245788
A.Accession: T03284
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1970 <ERI>
A.Cross-references: EMBL:AF007805; NID:g2367443; PID:g2367444
C.Genetic8:
A.Gene: mbc
A.Cross-references: FlyBase:FBgn015513
A.Map position: 3

```

```

Query Match 3.5%; Score 382; DB 2; Length 1970;
Best Local Similarity 19.0%; Pred. No.2,3e-13;
Matches 337; Conservative 286; Mismatches 596; Indels 553; Gaps 88;

Qy 500 SSKVAQVYLKNAKQACQLGQY--RMPAMARTL-----FKDASGULD-----KNAR 545
Db 320 SSRKASQLSVSSSGSSNGEYIIRPFGVACKDLPINKSDDRGNIDLPFIMCEKET 379
Qy 546 FSAIRQOSNLSNDMCKLADFRKPKMA-KLPIVILGNDITIDVNSDPEPNVNSY 604
Db 380 LDGTLRK--LIANDIKTI-----DSKMAVTVLEVLKGD---IKQIKKEPRLMHTN- 425
Qy 605 IPTKQFCSKPTIFVEBEFVPCIPKATQPYTI---YTNHLVY-----YPKYLKXDS 654
Db 426 -----VPYARKKGFBEVILPGVNRDLYLTITSGEFAARIAX-TS 463
Qy 655 QKSFARKANNIACIEFKQSDSEDSQPLKCIYRPGGVP/FTSAFAVALHHHON-DEFYDE 713

```

```

Db 464 EKQV---EVSVCVA---NEQGYLMPGVLSIGAGHPI---DEYKSVVYVYHDDKPKMOET 513
Qy 714 IKIEIPLTOAHKKNHLLTFEFHVSCDNSSGSTKCRDVEYQVSWLPLK-DGRVVTSE 772
Db 514 FKHINPIEDPFOCHLAFVLKH-----RSSNEQKORTEKPRGLAVRLMQANGTTITOG 566
Qy 773 QHI-----DVSNLPSGYLGYOELGMRHYGPEIKWVGSKP-----809
Db 567 QHILAVYKIDHKKYDKTVANNC-----YLEL-----PATVAELQAKPSIGGLTLPLPKD 614
Qy 810 LKISTHVLVSTYTOQHILHNFQYOCOKTESGQALGNLVKYLKSLHMEGVMTAFEP 869
Db 615 QLSIGVNLCSYKLTQSVSLGLLWMSAKET-----LEQSLNLSTVPGEEVYVFKQ 666
Qy 870 TILNOLFVILTRATQEE-----VAVNVTYVTHVVAQCHEEGLSEHLSRYKAVYKABEY 924
Db 667 DILDLAFILVENDHEKTDQVPMSTIH-LIETVSDLKHQPLSLVDVYINESFS-----721
Qy 925 VASEYKTHHEELTKSMITLIK-P---SADPLTSNKLKYSWFFEDVLIKMAOHLIENSKY 981
Db 722 FTLAATYKLMVLOKNISEAISPKESAD-----GNDLEESPEV 759
Qy 982 -KLNRQRPASYNHNAVETVYVNMMLPHITQKFRDNEBASIGNANSLAVTIKCFITMDG 1040
Db 760 RELYKITY--LHYVVKFVIRSRVLY-----AENCTDYMADATRLQELL-RM 805
Qy 1041 PVFKQINNYISCFAPGDKPTLFEYKFEPLRVVNCNHEHYIPLNPMFPFGKR-----I 1092
Db 806 FI-----DWIGC-----PSNLKSEGALK--NHIITATDLMQVBEHVRISISIVELL 851
Qy 1093 QRYYDQLDYS-----LTDFPCR-----N 1111
Db 852 EKFPRLTQSKMGKICKDFETGLFTLPKCRALLVPVCFKHIMHLESKEEIAECININW 911
Qy 1112 HFLVGLLAEVGTALOEPRERVLIAI-SYLNKLLHSGSDDRKASDAQRIATLYPLP 1170
Db 912 NTLKLLFRSDVSTNHNDIMILIRTYWK--AAHALD-----RTGLGKFF 958
Qy 1171 GLLIENVQRINRVDSPPVNAQMTYKDSLA-----LPAVNPLVTPQKSGT---1217
Db 959 AATMLGLQMDAQHYEYF-----VKDHOSELKHFYIEILVLEELVSPHQKAVFR 1011
Qy 1218 --LDNSLHKD--LLGAIIGSIASPYTTSTPNINSVRNADRGSLISTDSGNSLPERNSEKS 1273
Db 1012 DMWDMIMQNTVIGALKHL-----TVVID-----1037
Qy 1274 NSLDKQSGSTLGNSSVVRCDKLDQSEIKSLMCFYILYSMSDDALFTYMNKASTSELD 1333
Db 1038 -----YFLC--PEKQIWSNFPQCSIAFLVQ-----SPQLND 1068
Qy 1334 FFIISEVCLHOPQYMGKRYIARTGMMHARLQ--QIGSLNSLTFNHSYGHSDADY--LHQ 1389
Db 1069 FNDNKK--QOIVFAIYRDIRKDTAKETIKMWFQLO-----HKRKYFQVLYE 1112
Qy 1390 SLEENIATEVCLTALDLSLFTPLARKQLDLDHGNPLMKVFDVYLCPLO---KHOS 1445
Db 1113 PLEMSMIEKELRQ-ETIPIF-----FDMWQEBEYSSRHEHS 1150
Qy 1446 --ETALKVNFATLRSILYKPFYFEGRDMAALCYELKCKNSKLSIRTBASQLLYF 1503
Db 1151 YGDT-----KFNNAHHNG-----NFSDFKTAMIEKLDI 1178
Qy 1504 LMRNEDYTKGKSFVATHLOVIVSGLIA-DVVGIGTRFOOSLSIINNCAUSDILKH 1562
Db 1179 LIG-----AGKDAEIKHLEFETIMLRCAAHNTLVNDGAFVQVTRL-----MDKLLEY 1228
Qy 1563 TSPSSDVKOLYKRIATVLMATAQMKEHENDPEMLVDLYSLAKSYAST--PELAKTWLDS 1620
Db 1229 RPIIQO-----ESKENR-----MACTFSLLQFYSEVLDKEMYIRYVWK 1266
Qy 1621 MARIYQNGDLSAAMCY-VHTALVAEYLTRKEANQVQEPRLPHSHSACLRSRQGVVR 1679
Db 1267 LCALLHMEFENYTEAFTLKLHT-----ELLRMVDTLSHQ-----LRSYR---H 1307

```

```

Qy 1680 QGCTAFVITPNIDIEASIMEDVGNQDVHFNEDVLMELLBOCADGLWKAERYELADIYK 1739
Db 1308 NNCRRHRLQ-----KEALYFE-----IMEXFD-----XGKQMECALDMCR 1342
Qy 1740 LIIPYIEKR-RDPERLHMLYDTLHRAYSK-VTEVMSGRRLLGTYFRYAFFQQAAYOFT 1797
Db 1343 VLAQYEEBIFDYLLAEILNMLAFYEKIKELHNSE-----YFRVCFYGR-----1390
Qy 1798 DSEIVGEPFDEDEKXYIEKPLTPISEISQRLKLYSDKFGSENVMIQDSKANPK 1857
Db 1391 -----GPRPFLQNRVYIFRGEYERHSDFCARLVQHPQ--ALHQTEAPG---D 1436
Qy 1858 DL-DSKXAVIYQVTHYIPF-----FDEKELOERKTEBERSHNRFRFEMP--TQTK 1907
Db 1437 DITNSDQGYIQVKNKTEPIMGAFNFKNIINELVKKFTANNVQKQFSRFRSTNGC 1496
Qy 1908 RQGVBEQCKRRTITAIHCPEYVKRIIPWYQHHTDINPLEVALIDEMSKVABLRLC- 1966
Db 1497 DRDDVRNMLERTELRISYPLRPGILRWPEVETTFKISPLERAVEIMKDTNRDIQLVI 1556
Qy 1967 ---SSAEVDMITLOKLGSGSVQVNAQPLVAPAFLDOTNTKRYPDNK--VKLKVEPR 2021
Db 1557 LHKSETLHINPLSKLNGIVDPAWGGPAAKYEAPLTDDYUEQNPDPDEBELKELIA 1616
Qy 2022 QFVEACGALAVNERLIKEDQLE-YQEBMKANYREMAELSEIMHEQICPLE-EKTSYL- 2078
Db 1617 NOIPILDALIQH-RLPAPDSIKALQEHLERFAMQOQHEVRGYSKCDLKIENDSVVM 1675
Qy 2079 --PNSL--HIENALSGTPTSTVHGHTSSSS 2105
Db 1676 RRPNSFLPSLFDGNSNRHSET---SMGSSDS 1703

RESULT 7
BS3435
vesicular transport-associated repeat protein Tb-292 - Trypanosoma brucei
N.Alternate names: membrane-associated protein Tb-292
C.Species: Trypanosoma brucei
C.Date: 26-May-1995 #sequence_revision 07-Jul-1995 #text_change 07-May-1999
C.Accession: BS3435; S34395
R.Idee, M.G.S.; Russell, D.G.; D'Alessandro, P.A.; Van der Ploeg, L.H.T.
J. Biol. Chem. 269, 8408-8415, 1994
A.Title: Identification of membrane-associated proteins in Trypanosoma brucei encoding a
A.Reference number: A53435; MUID:94179225; PMID:8132566
A.Accession: BS3435
A.Status: nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-189,558-721,'P',723-738,'D',740-829,'L',831-1100,'Q',1102-1636,'L',1638-1648,'H',2420-2499,'L',2501-2550 <LEB>
A.Cross-references: EMBL:X73956
R.Idee, M.; Russell, D.; d'Alessandro, P.; van der Ploeg, L.
submitted to the EMBL Data Library, December 1992
A.Description: Identification of membrane associated proteins in Trypanosoma brucei enc
A.Reference number: S34394
A.Accession: S34395
A.Molecule type: mRNA
A.Residues: 1-2550 <LEB>
A.Cross-references: EMBL:X73956; NID:g393395; PID:g393396
C.Keywords: tandem repeat
F.74-557/Region: 8-residue repeats (A-R-L-R-A-E-E-E)

Query Match 2.6%; Score 279.5; DB 2; Length 2550;
Best Local Similarity 17.9%; Pred. No. 2.9e-07;
Matches 375; Conservative 306; Mismatches 758; Indels 655; Gaps 94;

Qy 245 RPAPELKM-----QDKSSYLAAADSE-VEMEMITLTKILQINFEAAMQEKRRGDS 295
Db 761 RPFENELMSPTVTYTTEDRNLPJHNSPDEQVLEBQGLRANQNVLV---AALSLRFTLGP 816
Qy 236 HEDDQSKLBGSGSLDYVLELAKSAREAEIKLSESRVLYFLYIDPDQOKLDFSSAEP 355
Db 817 VQVWEQA---CGFSLDP-VPGIRDGVRLPHTVYAGSSYIY---AHRIATNKPYP 866

```









1473 -----ERKEYLKESARNFSGQKPEKALAVYKDLKAYDEINYDNLGALFVHDQIAGIY 1527  
 1766 SKYTEVWHSGRLLGTFRVAFPGQAAQYQFTDSEPTDVEGFEDEDEKEIYVEPLTLP 1825  
 1528 TRLOST-----DRLVPTTFKVSFMG-----FGFPKSLNKSFFVEGLPEPHI 1569  
 1826 SEISORLLKLYSDK-----FGSENVKMIQDSGKVPKDLDSKYAYIOYTHVLP--PFDEKE 1879  
 1570 TSMHDLRLRSYHSGSNIVHSQEEVDML-----LKNP-----PMGKIYHVASVEPCISIDNYN 1621  
 1880 LQKRTKEFERSHRIKRFMEF--MPTQYKRGQ--GVEQCKRITLAIHCFPIYKKA 1934  
 1622 SSDKSSSI--NNVRYVYIENRDLRTFSNSRRLPGAGVTDLWVEEYTHMTMPTLMNR 1679  
 1935 IPWYOHHTDLPNIEVAIDEMSKVAE--LBOLCSAEVD-----MIKLOLKQSGSY 1984  
 1680 SEIVTKTKSLSPLENAIRSLQYKIOELVGLNCKMCKTLKDHDVNDLFTLESTNITGTI 1739  
 1985 SVQVNAAPLAYARAFIDTNTKRPYPNKVKLLKEVROFVEACGQALAVNERLIKEDOLE 2044  
 1740 SAVVNGISQY-KAFLEPSTSKQFSTDQLRLTARDELAVAVLGRCLTLHABELPEKDK 1798  
 2045 -----YQEMKANYRMAKELSEI----- 2063  
 1799 PSHDLLVRLFEENFAEIERYSRTLSBANNSRNMTARIISHKNPNKASFGSRDHHTS 1858  
 2064 --MHEQICPLEEKTSLVPLNSLIHFNAISGPTSTMYHGMTSS 2104  
 1859 GSNHSQFV-LEHSDSFGPNSL-LFGKYL--TTTSHSSTTS 1896

RESULT 9  
 A90551  
 conserved\_hypothetical\_protein MYPU\_3130 [imported] - Mycoplasma pulmonis (strain UAB CT  
 C:Species: Mycoplasma pulmonis  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
 C:Accession: A90551  
 R:Chandaud, I.; Helling, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
 A:Reference number: A9512; MID:21267165; PMID:11353084  
 A:Accession: A90551  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2819 <KUR>  
 A:Cross-references: GB:AL45566; PID:914089727; PIDN:CAC13486.1; GSPDS:GN00153  
 C:Genetic source: strain UAB CT1P  
 C:Genetic:  
 A:gene: MYPU\_3130  
 A:genetic code: SGC3

Query Match 2.1%; Score 225.5; DB 2; Length 2819;  
 Best Local Similarity 16.8%; Pred. No. 0.00045;  
 Matches 411; Conservative 399; Mismatches 851; Indels 781; Gaps 106;  
 15 KFTRA-LSKRGTAELRQSVSEVVRGSLVLAQ-PKLIPELDYENVVOKKOTIANDCLRE 72  
 621 KFTGATQNKILSHNSELVLRIVSGEKTFALEKIDDI-PNKAVYQKVEKIED----- 675  
 73 MLFPVDDDTQALIRQGRYICSTVPAKABEBAQSLPTECKTYSNMDHLVNYKEDYS 132  
 676 -FLFAVNTNEKAITDFVFLNELAKPE-----NFKALVSP 712  
 133 G--EFPQDLPNKV-----KLDELPHVVEVDEVDK----- 161  
 713 GLSQFLNNSNEVYSTQSAFYEVSKTLVAVDTSELKVLKNPIEIKSSAQKLNSP 772  
 162 -----DEDAASIGOKGKITKHGMLYKGNMNSAISVTWRSFKRRF-PHILQAGDSYRF 214  
 773 KERNLEKLOKQKSIISIFGFTNDLGMFLKQVYKSKLSSVEQDKTSSLKEVSVFVFS 832  
 215 EFLK-----DLQKEPKGISFLG-FLYGVSPFN-----NKVRRF----- 246

833 ENVKKLLISSACTYLSNIELFKE-NSDVFKGIFSHILNFEKTEPELLNKIIDVINSWESI 891  
 247 -----APE-----LKMQKSSYLLAADESEVMEEMITI-LKKIL-QLNFE--- 284  
 892 DOYNDAPFEPVNLFLKTNSSWFTYKVDIFSSVNSSEIQANILISKLNLSAQAGBELTN 951  
 285 AAMQEKRNDSHEDDEOSGLESGSLDSYLPBLKASAREAEIKLK--SESRVYKFLV 340  
 952 SSVESIKNIFTN-----TLKIADSNITDALVDTFVNSKTLLEIDKKNFIASIKNIQIFAL 1007  
 341 DED-----AQKLDSSAPPEVKSPEKKGKXI-----L 368  
 1008 SDDNFFSVFKALISSNKLKDKAFVNOYQODITNVLKEL-LFEKEFPNLIYSFMPKTSBL 1066  
 369 VKNCLSPNLOCCVANESEPTNVEPFFVTLSLFDIKYNRKISADPFVNDLHFSVRQMI 428  
 1067 FENNEBALNKTTIIRN-ALKKNLDPPIN-KMMDIIFD--NLDKIQLENYG--QVL 1118  
 429 ATTSBALMGSGPETQSALRGILHEAMQYPOKGIYSVTCBPHDIFLVARIEKVLQSGIT 488  
 1119 GLIKONSDALATKYQALFKOVVNEANFQDVLK-----VLVVSQTKKFN-GPEL 1166  
 489 HCAEPYMKSSDSSKYAQVYLN--AK-----QACQL----- 518  
 1167 NADETNKISTDLTKIKOVFESNFKSLALATKQSLVIDGVQANKIIFDPTPLNTFL 1226  
 519 -----GOYMPFAMARTLFKDASGNLDKARFSAIYRDSNKLNDMDL-KLADF 569  
 1227 SEKDEKENTQMVMS-ILNTKLIDISQNSSEIKRLTLFTKSA--LSRDEYETKIVF 1283  
 570 RKPERAK-LPV-----ILGN-----LDITDVSDEPNY--VNS----- 602  
 1284 LSKEKLASVTPVEFVKFLNKEKTNFSDDLIDSLPFDFOQOANSFEELIYVLRN 1343  
 603 -----STLPTQFETCKPTTFEVEFVPI--PKHTQPI-----TIYNHLYV 645  
 1344 NKEIVNVYLKELSEFIKSNITIQRLGOTINALVTPNEKIFSDTELQSLSLINESDL 1403  
 646 YPK-----YLKVDQSKFPAKARNIACIEFKDSDEDS--QPLKCIYGRP 688  
 1404 FCKTIIISNLDVLUSALINQESSVLTKEVNT--HKAIDQNSYFAVIKILLSVK 1459  
 689 GGPVETRSAPFAVLHNNQNPFEYDEIKLEPTOLHEKHLHLLTFP-----HVSQDSS 741  
 1460 EGDVDKSDPEKAIKKPLN-----LEAKDVIYLTKEIISPVDKTIQDFDSMT 1505  
 742 KQSTKRDVETQVQYISWLPILKQGRVYTSQHIIVSANL-----PSGLYQELGNG 794  
 1506 KSIIVK-----LIKV--VLSDBONLKFSAQIITDIFANKRYLAVENLN-- 1547  
 795 RHYGPEI--KWVDGKPELKLKISTHLVSTVYTOPDHLNFPQYCOKT--ESGQALGNEL 849  
 1548 -----DFLIDALVDKEMINIUSNTLGKITDEPKKIISFIRKNMTSBSSTTEBI 1602  
 850 VKYLSLHAMEGHVIA-FLPTIINQLPRVILTRAQEBVAVNVTVIIVVAQCHEGLE 908  
 1603 NTIVASAEHLISYVDSNFMQLKDFALPEKKNKQDSKTSLE-----FS 1648  
 909 SHLRKYVKAIAEYVASEYKTV-----HEBLTSMITILKPSADFLTSNKL 957  
 1649 TFMKSISEBFLDAENNYALFRKVLSSNLSSTOSDQSFYTSISSILIRV--FLKEKAL 1705  
 958 KYSWEPFVLLKSMNQHLIENSKVLLNQNQFPASVYHNAVEVVMMLPHITQKRDPE 1017  
 1706 -----ELMLK-----KQFNPEL--HDIKVDADNL--IKFIFRN-- 1737  
 1018 ASKNANSLAIVFIKCTFMDRGFVFOINNYISCFAPGDPKTLFEYKFEELRVVCNHH 1077  
 1738 -----QVIDPI-----ETIIN-----RVVSQKEE 1756  
 1078 YIPLNL-----PMFGKRIQRYODLDYSLTDEFCNNHPLVGLLAEV 1122

Db 1757 YKLSNYSOELIYKPLSSNKNETLNFPKQIGETKNSKLINPLAKSFISEB-----LKKS 1810  
 Qy 1123 GTALQEFREVRILAI SVLKNLILKHSFD--DRYASRSHQARIATATLVLPFGLLIENVORI 1180  
 Db 1811 KTLTON-DEIDSI VNFVNSLVLKTDLID-----KLYDLVSSFPNAKVSPTKTV 1860  
 Qy 1181 NVRDVSPF--PVNAGMTVXESLA---LPVAV-PLVTQKSGSTLDSLHKDLGAIISGI 1233  
 Db 1861 EIFNFSLSLKPI---MDQKNYFSLVKTATLITNMKLMESQNTDALKNALGQMLLRVFK-- 1915  
 Qy 1233 ASPYTTSTPNINSVNRADSRGSLISTDSGNSLPERNSEKNSLDKHQOSTLGNSVVRCD 1293  
 Db 1916 -----NLDLNLKFLDPVLQNIINFLDQOKSKTKENLNR-- 1949  
 Qy 1294 KLDQSEIKSLMCFLY-----ILKMSDDALFTYWNKASTSELMDFPTISEVCLHQFOY 1347  
 Db 1950 ---NASKQILVFLVLONEOTLEPFKSLVNNIISNNQNYGANNYLDLY---SFIENKQ 2002  
 Qy 1348 MGRYIARTGMMHARLQOLGSLDNLSTFNHSGHSDADVIHQSLLEANIATEVCLTALDT 1407  
 Db 2003 LVTQYFKKIAPKIASDSMLKTFVNGFALNLQEQN---LBSKDELISVNSITSGAIDA 2058  
 Qy 1408 LSLFLTAFKQULADHGNPLMKKVVDVYLCLQKHQSEFALKNVFTALRSLLYKFPSTF 1467  
 Db 2059 LST-----SNIVEKILEVFLAEVKKQNPQSKTNDLPASIQ----- 2093  
 Qy 1466 YEGRADMCALCYEILKCCNSKLSIRTEASOLLYELMENPDYTKSGSFVRTHLQVITS 1527  
 Db 2094 -----ELIK--SKDFSQNGLFVLSIILKSN----- 2118  
 Qy 1528 VSQIADVVGIGTRPOQSLIINNCSNDRLIKHSPSSDVK-----LTKRI----- 1576  
 Db 2119 ---ILDSKINNKEYIQVKYTL---PEKLIKNTSVQNFVKDKETNISISPIEQNN 2169  
 Qy 1577 ---RTVLMATQMKHEHNDPEMLVDLQYSLAKSYASTPELRK-----TWL 1618  
 Db 2170 KQSLSKIVALLASDPOTNLVKNKVIDIDIFKNHGVNASKANLEIILAEFLTTHKQDLAYL 2229  
 Qy 1619 DSMA-RHYKNGDLSAAMCYVHTALVAEYLTR-----KEAVQWE----- 1658  
 Db 2230 KATINKLVKSELHDFP-----VNVLLSQLSKNIIFENNEKEQKSLKTSIHLIIN 2284  
 Qy 1659 ---PLLPHSHSACLRSSRGVFRQCTAFRVITPNIIDEASME-----DVGQDVH 1708  
 Db 2285 SDIPQVANNLJBELNGKK--ALSRSVSQKITTRLNLPDKGFELFKVLANSVQKQNN 2342  
 Qy 1709 FREDVLMELLEQCADGLMKAREYELIADYKLIIPYERKRPBERLAHYDTL--HRAYS 1766  
 Db 2343 FROEVINQIKNGLSIYLNKDTLINALFTNFDLKIPNDLQNSFNLKEISNDVHFHRYE 2402  
 Qy 1767 KTEVWHSGRLLGTFRVAFPGQAQOFTDSETPVBEJFDEDEQKEYIYKPKLTPIS 1826  
 Db 2403 AILLEVFSQ-----IHNNGKIYSAQKSEIKQW--EDLIRVLKKSQSVANN 2450  
 Qy 1827 EISORLLKLYSDKFSSENVKMIQDSGVNPKDLSKRAYIQVTHVLPFDEKELQERKTE 1886  
 Db 2451 KYEDALYDLI-----YKLLKSNLSK-----NILVSQLPMF-----XSAKD 2488  
 Qy 1887 FERSNIRFRMFBMPFTQGRKQGVBEQCKRRTILTAIHCPYVKKRIPIVYQHHTDIN 1946  
 Db 2489 IYTKHHEILIRFOQIGRAHEQG----- 2512  
 Qy 1947 PLEVALIDEMSKYVABLRQCSAEVDMITLQKLQSGSVGVNAGPLAVARAFLDPTNK 2006  
 Db 2513 ---MLKNI FKSLSLDFKNTNFSKVPFNKLTLEIMNIYSN-----FMDKXTNT 2557  
 Qy 2007 RYPDN--KYK-LKEVFRQFVEACGQALAVNERLIEQL-----EYQEMKANRYENA 2057  
 Db 2558 LVVSNIMTKDLDKVFRTYRDAKQFVNFPMNLFFASLSLPKNTMNGYQ-----NGYI 2610  
 Qy 2058 KSLSEIMHOICPLEKTSVLPNSLIHFNALISCTPTSTWNG 2099  
 Db 2611 KMLTNTITY-----PNEANAGA-QGSSSLKAYDG 2638

RESULT 10  
 T08611  
 hypothetical protein DocA - slime mold (Dictyostelium discoideum)  
 C/Species: Dictyostelium discoideum  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 C/Accession: T08611  
 R/Auby, L.; Firtel, R.A.; Iranfar, N.  
 submitted to the EMBL Data Library, August 1997  
 A/Reference number: Z16456  
 A/Accession: T08611  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-831 <AUB>  
 A/Cross-references: EMBL:AF020409; NID:g2425146; PID:g2425147  
 A/Experimental source: strain AX4  
 A/Genes: docA  
  
 Query Match 2.1%; Score 224.5; DB 2; Length 831;  
 Best Local Similarity 17.4%; Pred. No. 7.3e-05;  
 Matches 158; Conservative 143; Mismatches 297; Indels 311; Gaps 35;  
  
 Qy 1234 ASPYTTSTPNINSVNRADSRGSLISTDSGNSLPERNSEKNSLDKHQOSTLGNSVVRCD 1293  
 Db 140 AATATATSTLTNNNNNSNNNSITITNNNN--NNNNNNNNNDIGSSAYTGSGNSLI--- 195  
 Qy 1294 KLDQSEIKSLMCFLYILKMSDDALFTYWNKASTSELMDFPTISEVCLH---OFQW- 1348  
 Db 196 --NQSQF-----FTQWR-----SFLTLTSYVINCQDLQFESVN 227  
 Qy 1349 -GKRYIARTGMMHARLQO-----LG-----SLDNLSPNSY 1379  
 Db 228 PAKAVFITRCGDVAIEARPERVWATVPYKERSTFVLYGLPIVKLSISTPMKRYA 287  
 Qy 1380 GHSDDVLIHQSLLEANIATEVCLTALDTLSFTLAPKQULADHGNPLMKKVVDVYLCLF 1439  
 Db 288 TFIPTDMESEILQNGTDTLFYHTIDSL---LEISTERL-NRGWPIVCRSVGDKLPF 342  
 Qy 1440 LQKHQSEFALKNVFTALRSLLYKFPSTYEGRADMCALCYEILKCCNSKLSIRTEASQ 1499  
 Db 343 -----TLKSP-----STEFNKK--CWSL-----VLKSTDAITKSEQ 373  
 Qy 1500 LLYFLMRNPDYTKGKSPVTRTHLQVIVSQIADVVGIGTRPOQSLIINNCSNDRL 1559  
 Db 374 FTHDI--NHFLY-----L 384  
 Qy 1560 IGHTSFSSDVK--DLTKRIRFTVLMATAQMKHEHNDPEMLVDLQYSIAKSYASTPELRKTW 1617  
 Db 385 I--PFPNDVVSQDEBELFSSVSKLLYVLQEHKRTAHFI-----RF 423  
 Qy 1618 LDSMARIVKNGDLSAAMCYVHTALVAEYITRKAQVQWEPRLPHSHSACLRSSRGV 1677  
 Db 424 VMTSRRRHYEMGNYTEAA---VTLMHLASL---YQWD----- 454  
 Qy 1678 FROGCTAFRVITPNIIDEASNMEDVGMQDVHFNEDVLMELLEQCADGLMKAREYELADI 1737  
 Db 455 ---ANKVIAQTAASEIGSFPEQK-----ESRKEKL--L 481  
 Qy 1738 YKLIPIYEKRDPERLAHLY-DTLHR-----AVSKTEVWHSGR 1776  
 Db 482 YKEILLCTNNGAMARALPIYKELVHNHTNMIQDMSAATYLRQGAFFQKINE---SAD 538  
 Qy 1777 RLLGTFRVAFPGQAQOFTDSETPVBEJFDEDEQKEYIYKPKLTPLSISORLLKLY 1836  
 Db 539 PVFEDYFVGVYGG-----KGFSSIONKEFIYKGNQFDRLSDFISKI----- 580  
 Qy 1837 SKPFSENVKMIQDSGVNPKDLSKRAYIQVTHI-----PFPPEKELQ 1881  
 Db 581 QKMKPSELKLTTEVPQOIQDSQOYLLITSVNSINIGFQKRHDSANASITNKGRVP 640  
 Qy 1882 ERKTEFERSHNIRFEMPFPOTG---KROGVBEQCKRRTILTAIHCPYVKKRIPIV 1938

Db 641 HRVQGFNARNKVVNFVYKRFKKNATATATSONEFEDLWMNLYFCVENSEFPCRCLIV 700  
Qy 1939 YQHTDLPNLEVIDEMSKVVALRQLCSGAYD-----MITQLKQSVSVQVN 1989  
Db 701 ERKQVHSPLNENLNSIVQKNDL-----AARIDRHOANTOESISPLTMLNLTIDASVN 755  
Qy 1990 AGPLAARAFLLDDTNRKYPD--NKVLLKVEPRQFVEACGQALAVNERLIKEDQLEYO 2047  
Db 756 GGVSRF-ETFMSEDYIKQNPVEYKNIAELLKXSALDQQLVIVEQGLKXHSQLRPEMAAQD 814  
Qy 2048 ENKANTREM 2056  
Db 815 KLETFEFTM 823  
RESULT 11  
G70163  
Hypothetical protein BB0512 - Lyme disease Spirochete  
C:Species: Borrelia burgdorferi (Lyme disease Spirochete)  
C:Date: 13-Feb-1998 #sequence\_rev: 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: G70163  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horec, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A>Title: Genomic sequence of a Lyme disease Spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; PMID:98065943; PMID:9403685  
A:Accession: G70163  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2166 <KB>  
A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PIDN:AA06876.1; PID:G268842  
A:Experimental source: strain B31  
Query Match 2.0%; Score 222; DB 2; Length 2166;  
Best Local Similarity 17.4%; Pred. No. 0.00047;  
Matches 407; Conservative 378; Mismatches 800; Indels 752; Gaps 105;  
Qy 15 KTRFASRSGTAAELRQSVSEVVRGSLAKPKLIPLDYENVIVQKTIQLNDCLREML 74  
Db 26 KGSRALDK-----IKKFVDLTKNLEP-----FIEDKKEINDLAVDM- 63  
Qy 75 LPEYDDPOTAILRQGRVYICSTVPAKAEBAQSLFVTECIKTNSDMHLVNYKYEDYSGE 134  
Db 64 ----EAYQSSIS-----IIRKIE-----VQKIKKNSDPAVEKKIAVHDSM 104  
Qy 135 FROLPNKVVK-----LDKLPVHYVEVDEVDKEDDAASLSOKGKITKGM 180  
Db 105 LKDLDEMTFQVQNIQRLQVQDKIVDKLSTKLGFTQIDSYE--SNLSV--LEKFDK 159  
Qy 181 LKGNMNSAISVTRKFRFHLIQLGDSYKFEFLKDLQKPKSIFLGFLYGVSPFN 240  
Db 160 ANKENMESIKIASWEKFDNIKELV-----FKID--NLNKEIS-----LYEDDLAN 203  
Qy 241 NKVRFAPELKQDK-----SSYLLAADSEV-----EMEWITILN--KIQLNFEAM 287  
Db 204 IERKNDILVKNKEDLDESPLEKVEFNIGKYSKEISSFFNYNKKYKLITNSJELIM 263  
Qy 288 QEKRNQDSHEDD-----EQSKLEGSGSLDYLPELAKSAREALIKLSESRVLFYLDP 342  
Db 264 ESKVKNINEKEDFILRLNBELOKFKDILVVD-----RSKEIKDLKEDKLV- 313  
Qy 343 DAOKLDPSSAEPVPSFEKFGKRIIVKCNDSLFINQCCVAENBEQPTTNVEPFF----- 397  
Db 314 -----VDNEISSSSSFFKNVYSRINSLEESWKIEGKYE-----QVDDVPDKFRS 360  
Qy 398 -VTLSTLFDI--KNRRTISADFHVDLNHFSRGOMIATSPALMNGSPETQSAL----- 447  
Db 361 QVEBLNKATYEDYEDKIS--QVDNN--IREKV-ELSLDLNLSKNESVQSGAIDFIKEL 413  
Qy 448 ----RGILHEAMQVPEKQ-GIFSVTCPHPDIFLVARIEKVLQGSITHCAPYKSSDSK 502

Db 414 EDDNSGIYILEFGKRGADIEVSESFKGDINDQLKQLESQLLDVDSNIOEKLIKLNND-- 471  
Qy 503 VAQKVLKAKAKACQCLGQYRMFPMAARTLFPDASGINDK--NAFSAIYRQDSKLSND 560  
Db 472 ----LISFEEINGR-----FNNNSYLNNDINAKYALFE----- 503  
Qy 561 DMLKLLADFRKBEKAKLPLVILGNLITIDVNSDPFVYVNSY----- 604  
Db 504 -----SLDSSSSKFEENQMEKYSFIDKYLACMDERSL 536  
Qy 605 IPTQPEFSKTPITFEVEBEVPCIPKHTQPTIYNHLVYPRKYLKIDQSKSPAK---- 660  
Db 537 MYGEKFETLSQ-EATNNYQEFQDLNKLKLENEIESFYNNFEKTOETLKYDFNTSLINKE 595  
Qy 661 -ARNIAICERPKDSDEDSQPLKCIYGRPGGVFRSAFAVLHNNQPEFDEIKIELP 719  
Db 596 IGRNI--VEFRD-----RYDEVNI-FV 615  
Qy 720 TOLHEKHLILTFPVSQDNSSKSTKRDVETQV-----YSMLPLKDGAVVTS- 772  
Db 616 TOL-BESKLOYSKMGEMDSNKN-----IESQINTNEBFLSLIQDKXGJELSES 667  
Qy 773 -----QHI PVSANLPSGYLYQELGMRHRYGPEIKWVGKPLIKI--STHLVST--- 820  
Db 668 VFNDSLDIH-----QKKAIDHSGMKDELIALNKSLLIDIKVSEBLLSATL 714  
Qy 821 -VYTDQHLNHFQYQCKTESQAQALGNEILVYKLSLAMEHVNVIATPLTINLOLFRVL 879  
Db 715 KESLEKQVNDMEVYLLKKTGIESLVEKKEKELKMSYSQSDAIIIGIKEFINQTEII 774  
Qy 880 TRATQEEVAV-----NTRVILIHVAQCHE-----GLESHLRGV 915  
Db 775 -----KDSVPELLEDLNNKFPDKNNFVLSKIECDYKLDKPIESBIDLNNKSDLNBEI 829  
Qy 916 KYAAYAEPPVASE-YKTVHEELTKSMITIL--KPSADPLTSNKLKYSMPFPDVLKSM 972  
Db 830 ESKLQIVSNIKSDNQIIDFLDRISKDILNRKOSINNEVSKLSQSDWKLSNLEITVK--I 887  
Qy 973 QHLINSVYKILNRNRPASVYHNAVETVYNNMLPHIITQKFRNPASGNNHSL-AVFIK 1031  
Db 888 ENLLSGKVDL-----DLIDSEVTTIKIE--LPSIESLSYLYE 925  
Qy 1032 RCFETMDRGFVFKQINNYSICFAPGPKTLFEYKKEFLRVVNCNHEYIPLNLPFGKGR 1091  
Db 926 KIDFERNGAI-----YSBELLQDINMH-----FNKET 953  
Qy 1092 IQRYODLQDYSL-----TDFCRNHLVGLIREVGAQLQEFR--EVLIAISYVKULL 1144  
Db 954 RELBENLSKKEFAVLNNESEEFYKE--VDSLQDKRTDIAFQANIDITLDSLVNKFNDI 1010  
Qy 1145 IK-----HSFDDRASR-----SHQ-----ARATLVLPLFGLLIEVQVRIN 1181  
Db 1011 NKEINGKYNVISNYRGYSENISSKLENEIMHEINLSRRLRIDRIDSLSKGDENLQK 1070  
Qy 1182 VR-DVSPPEVNA-GMTVXD-----ESLALPAVPLVTP-----QKSGTLNDS 1221  
Db 1071 EBFQVSKYQVEKELKVDLDDGEA--KINKLVKIEQYKRSLEALDYRRTIND 1126  
Qy 1222 L--HKOLLGAI-----SGIASPYTT-----STPNINSVR 1248  
Db 1127 IMQAKERFGEITNELNKNIESKSEFLNDLYKERFYLIESNFEERYSTFLIESEGASIKR 1186  
Qy 1249 NADSRGSLISTDSGSLPERNSEKNSJDKHQOOSTLGNVYRCMDKQDSEIKSLMCEL 1308  
Db 1187 D-EIYKTLTNSD--ENLDIKISEMDONFEIIEORS--KDLIEFEKLODKTKD--CYG 1237  
Qy 1309 YLKSMSDALPTYNKAQSTSE-LMDPTISEVCLHOFQYMGKRYIAR-TGMGHARLOOL 1366  
Db 1238 PINSQGE-----IKAGVEINIGNH--DVCIKAVTLIDDIIVKENEIHKRISL 1287  
Qy 1367 GSLDNSL-----TFNHSYGHSDADVLHQSL-LEANIATEVCLTAL 1405

Db 1288 KSIESTPPSIEKNLDKVGSCIDKIANDFNLKYIELEERCNQGLNENKIDNKI--KAI 1345  
Qy 1406 DTLSFTLLAFKQQLADHGNHPLMKKVPVLYLCLQKHOSFTALKVFTALSLIYKPPS 1465  
Db 1346 DNLAL-----SOYDLKKYADMTDEFSSR-----LMSYATLASE 1380  
Qy 1466 TFYEGRADWCAALCYEILKCCNSKLSIRTEASQLLYFLMKNPFYTGKKSFWRTHLQVY 1525  
Db 1381 EFKSNKEMIFEL-----ESQLKNLKNLESDL-----NNVE----- 1411  
Qy 1536 ISVSQILADYVGIIGTRFQOSISIIINCANSRLIKHNSFSDSVDYDLTKRITVLMATQ 1585  
Db 1412 -----KDVIRLKEESY-----HNVSMLKLEEDPF-----KDLKRGELKYSLEN 1453  
Qy 1586 MKEHNDPEMLVDLOYSLAKSYASPELAKTLDMSARHVKNGDLSAAMCYVHTLV 1645  
Db 1454 FIASIND--KIONLEYSKULENTLQSF-----RLDIOKAKDKENYTLDPF--- 1503  
Qy 1646 AEYLTRKAVQWEPPLPHSHSACLRSRGVFRQCTAFRVITPNIDEASMMEDVGMQ 1705  
Db 1504 KEFSSKKCKMOSEIALMETN-----ITGVDEFVDF----- 1534  
Qy 1706 DVHFDVIMELLEQCADGL--WKABRYELIA-----DIYKLIPIYEKR 1748  
Db 1535 --VNNKOSIIDSWPLNIKQDVKMOEKSYSTIEKRINLAELGIKSPENDIFNVYKIGLESFK 1593  
Qy 1749 RDPF--RLAILYDTLHRAVSKYTEVMHSGRLIGTYFRVAFPGQAQVQPTSETVEGPF 1807  
Db 1594 DGFELKAEIIFSNLQNEAKKIQOSVHLDPKNIGESLNKYL-----DEKTV 1640  
Qy 1808 EDEDEKEYLYKEPKLTPLEISORLKLYSDFRGENVMQIDSGKVPKQDLSKYAVIQ 1867  
Db 1641 D-----FK-----LEKIDKVNKKTEDILIQAEVFL-----TQCKDLEK----- 1676  
Qy 1868 VTHVYFPFBEKLOERKTFERSHNRPMFEMPTQIGIRQGVCEBQK----- 1917  
Db 1677 ----IFELQKLEHETTLSSMLDKVRREMVVISDSKESFEGQIETLKNKISERSEKIS 1732  
Qy 1918 --RTILTLNI--HCFPVYKRIIPVWYQHNHDLNPIEVALIDMSKQVLAELQLSAEVDM 1974  
Db 1733 LYRNNIETSIENYNSFSKSI-----KSDLGLLE--DHLKSKLKH-----STSIETI 1778  
Qy 1975 KLQKLQGSVQVNAAPLADNTKRYDPNKKVILKEVFRQFVEACGQALAVN 2034  
Db 1779 KSGLQEQ-----IDKEVEYFPKKNHKKELKEVDNNMLLEESKILIND 1819  
Qy 2035 ERL-----IKEDQLEYOEEMKANTREMAKELSEIMHEQICPLEKTSVLPNSLHP 2085  
Db 1820 VQFNKFISEIKDNLVEYKSLDRAEFEDSYDKINFOIENQIENFKKIDSELEKNSIF 1876

## RESULT 12

T28676  
rhotptry protein - Plasmodium yoelii (fragment)  
CSpecies: Plasmodium yoelii  
C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 01-Dec-2000  
C.Accession: T28676; A45521  
R.Stuba, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A.Title: Comparison of two members of a multigene family coding for high-molecular mass  
A.Reference number: Z20507; PMID:97077455; PMID:8920022  
A.Accession: T28676  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-2401 <SIN>  
A.Cross-References: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:AAB41263.1  
R.Keen, J., Holder, A., Playfair, J., Lockyer, M., Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A.Title: Identification of the gene for a Plasmodium yoelii rhotptry protein. Multiple c  
A.Accession: A45521; PMID:91101660; PMID:2270106  
A.Status: preliminary  
A.Molecule type: DNA

A.Residues: 2260-2401 <KE>  
A.Cross-References: GB:M34281

Query Match 2.0%; Score 214; DB 2; Length 2401;

Beet Local Similarity 18.1%; Pred. No. 0.0016;

Matches 385; Conservative 336; Mismatches 760; Indels 650; Gaps 104;

Qy 54 YENVIVQKTOI-----LND---CLREMLLPYDDFQTAIRRQGR-----YICST 96  
Db 597 FENILTDKKTLETFETPOLSLNNHNNKELLTYFD-----LKNLJKNENMLYKQFN 651  
Qy 97 VPAKAESE-----AQLPVTECIKTYNSDMHLYNKYEDYSSEFQLPNKVYL 145  
Db 652 EKEKAVEDIKKKNVDINKIVANIBITVTSLYN-----INDETEIASKSIELANTKV-L 705  
Qy 146 DKLPHVHYEVD--EVDDEDAASLGSGKGIITHGMVYK-----GNNNSAIS--VYMR 195  
Db 706 EKVKANVTNLNIEIKELKDYPDFQPGKEKN--IKYDENKIKNDI DTLNOKIKDSIEFTLT 763  
Qy 196 SFKRRFFHLIQLDGSYKFEFLKDLQKEPKSIFLGFYGVSPR--NNKYRRPAPFLKMQD 254  
Db 764 EIKKSEHNDIEIQ-----IDKLKVPNKTMTFREDPEKEIKKENIVKIKDKKNVYK 818  
Qy 255 KSSYLLAADSEVEBEMWITLTKILQLNFEAMQEKRGD--SHEDDQSKLEGGSGGL 311  
Db 819 EIDKLINISIKENDK--TSLEKLNINLSYG--KSLGNFLQOIEEKKKAETHIRYAM 873  
Qy 312 DSYLPEL--AKSAREAF-----ITLKSERYKLYFDLPDPAQKDPSSABPVKSPF 360  
Db 874 EAYIDDLNIRKKQOEIEKENIMMDIKMDIHKEKKNALNISHDVKIYHTTS---KNHE 929  
Qy 361 EK-----FGKRIIVKCNDSFNLQCCVVAENEGAPT-----TNVEPPVYL 400  
Db 930 EKISDIRKNSLKIIDPFSEBSYI--NDIKKELEKRVLESQNNNDINDIYLSKINIVYL 987  
Qy 401 SLFDI-----YNNKISADFHVDLNFHSVROMIATSPALNNGSGPETQS 445  
Db 988 KLNKIKKIIDKVEYTDIEIKNNKINAE-----LSNSEKITQLE--NSLIKQCS 1038  
Qy 446 ALRGILHAAMQYPRQGISFTYTCRHPDIFLYARIEKVLQOSITHCAPYMMSSDSKQAQ 505  
Db 1039 KIKSTIID--NYSECIKNIT--NLKTYIYENKNI-----HTFYKAAEYNOV 1084  
Qy 506 KVLKNAQACORLQGYRMPFAMAARTLFPKASGNLDKARPSATYRQDSNKLSDNM-LK 564  
Db 1085 SLNPNINIMADTKGQYIL-----NIKKN-----NGTNNTDVNIK 1118  
Qy 565 LLADEFKPEKNAKLPVILGNLIDITIDVNSSDPPYVNVSSYIPTQPFITCSKTPITFEVB 624  
Db 1119 ELKEHKKKSNNYK-----DEAGKN-----TQELIKK 1143  
Qy 625 FVPCIPKATQPYTYITNHLVYYPKYLKDSQSKAPKARNIACIEFKDS-----D 674  
Db 1144 NKELEFKEQEVYVLLNKYVAVELKNKEDKTKNYS-----QIIRKIDAHNTFTSQADKS 1199  
Qy 675 EEDSOPLCICVPRGGPVP--TRSAFAVLIHNOHPEVD----- 712  
Db 1200 EKKKNEIENQIRLEDVAVANKKNSKAILDQLSVPPKIRFLKIKDLRTSDDLCKRTK 1259  
Qy 713 --EIKIE-----LPTOLHEKHLLLTFPHV--SCDSSSKG--STKRDRVETQV--- 755  
Db 1260 DIETKISMLSIDQETKLKENKILNTLEKLSIKQKKNIDQKKELDEVNSKIKNIE 1319  
Qy 756 -----GYMWLPL-----KQGRVYTSQHPVAVANLPSSG----- 785  
Db 1320 SNNVQHKKNYB--IGIVEKINIRAKNQDIESTQKLIPIIKNLISPKANDLEGIDITNK 1378  
Qy 786 -LGVOELMGHRYGEIKVNVGGRPLKISTHLSVTVYTDQHLNFPQOYCKTESGAQA 844  
Db 1379 NLGKNTNEMNNIYEFIKSYD-----LTHVLETYSKEP---ITYEQIKKKRITQD- 1426  
Qy 845 LGNELVKYLLKS-----LHAMEGHVMIAPFLITLNLQFRVLTATQDEAVVNV----- 891

Db 1427 --NELLTNKNVAKSYLDIEANEFDRIVTHFNKRLNDVNDKFTNEYSKVNGPDNIS 1484  
 Qy 892 -----TRVTHVVAQCHEEGLSHLSRYVAVAKAE----- 922  
 Db 1485 NSINNVKSTDENLNLINLQTKEMANTVSKYYSKYAEABEIPINIPKLANSLNIQIK 1544  
 Qy 923 -----PYVASEKYTVHEELTKSMITILKPSADFLTSNKLKXSWFFPDV 966  
 Db 1545 SSSGIDLPKNINAIPLDLSQCK-----DLTFTIPSEPKSEYTYTMSOSYNTLLDI 1597  
 Qy 967 LKSMAGHLIENSKVLLRNQRPASVYHNAVETVVMMLPHITQKFRDNPESKANSHL 1026  
 Db 1598 LKRSOELQKKEQOALWLEFENRL---LHOKVQA-----TNEKIDTLSDLKXKEQEI 1645  
 Qy 1027 AVAIFKCTFMQGFVFKOIN--NYISCFAPGDPKTLFEYKPEFLVNCHEHYPLNP 1084  
 Db 1646 LNKVK-----LLHKSNELNKLSGNSONVDTLSESKDKIKESKNNEKEKENIG 1696  
 Qy 1085 MPFGKRIQ-----RYQDLQD-YSLTDFPCRNHPLVG-LLREVGTAQ 1127  
 Db 1697 INFDVAKMEQFNNDIKDIKLENNYKSEKNVNSE--NNNLSQSKKLKELTNAN 1754  
 Qy 1128 EREVEVLIASV--KULLIKHSFDDRYASRSHQARLATYLPFLGLLENVORINRVDS 1186  
 Db 1755 --AEIKKIEDKIEKNGKLNKLIEFR-----KDCMLFYKTLVETL-KIKTTDYT 1801  
 Qy 1187 PFPVNAGMTVYKDESLAPVNPVLPVTPQKSTLDNLSLHKLGLASGIAPIYTSTINIS 1246  
 Db 1802 KFTTSTKRSKELKTIADTS-----NSLNDP--INLTQTKY-----DLNQ 1840  
 Qy 1247 VKNADSRGLISTDSGNSLPERNSEKSLDKHQOOSTLGANSVRCXDKLDOSEIKSLMC 1306  
 Db 1841 INGHVSWVADATNDNNNLEKEKEATKTNLNTLFTIDSKNIDAGLHNNKIQI 1897  
 Qy 1307 FLYILKMSDDDLFTYWKASTSELMDFFTISVCLHQQYMGKRYATGMMH-RLOQ 1365  
 Db 1898 -----YFN-----SEL-----HKSIDSIKOLYK-----MHAFLKLN 1924  
 Qy 1366 LGSLSNLTFNHSGYSHDADVLHQSLEANIATEVCLTALDPLSLTTLAFKQGLADHGH 1425  
 Db 1925 IGH1-NKKYFDISKEDNLTLOLESILTANL-----NDL----- 1957  
 Qy 1426 NPLMKRVFVYLLCFLQHQOSETLKNVFTALSLEYKPSFTEGRADCAALCYEILKC 1485  
 Db 1958 KEIGKIDSKKKQFLHA-LSEFTIPN-FMTLKEIHYDIK-YKQID-----ELENI 2006  
 Qy 1486 CNGKSSIRTEASQLLYFLMRN-----NF--DYTGKGSVTRTHLQ-----VIISQSLIA 1533  
 Db 2007 TNEENENI-TLYMDITTKLMKVKQSIINFTVYENDSNITIKOHIDNNENDVSKIKESIE 2065  
 Qy 1534 DVVGIGTFRQOGLSIINCA-----NSDLIKHTSPSSDVKDLTKRIFVLMATAQM 1586  
 Db 2066 TTI-----QSFQKLKLNLEIKAFYNNNNINVT--STISQDVNDVKHISDGLT----- 2115  
 Qy 1587 KEHENDPEMLVDLQVSLAKSVASTPELRKTWDSMAR-IH-----VKNGDLEAM 1636  
 Db 2116 ---BNE---LIGQKLEDDIKKSTVDIRGEQITKYVNPJHDVYEQTKIQQNNPKMD--- 2166  
 Qy 1637 CYAVHTALVABYIT-RKEAVQNEPPLPSPHSHACLRSGVGFROGCTAFRYITP--NI 1692  
 Db 2167 ---EIDDLLOEIVNVKSESELKPLTIINKDN-----VTPPIISRI 2203  
 Qy 1693 DEBASWMEVDVQMODVFNEDV---LMELLEQCADGLMK--ABRYELLADI---YLLIIP 1743  
 Db 2204 DKYINLIK---SEYNNNDVSVYNVAKKLEEDANNIIRLDTSNNLNDLLOQNFKITID 2259  
 Qy 1744 IYKRRDPERLALVDTLHRAVSKVTEVHNSGRRLIGTFRYAVAFGQAQOYFTDSETDY 1803  
 Db 2260 LKKKQGEIENRNML-QTINRE-QEIQOTEH---VANNYTH---DINDI 2299  
 Qy 1804 EGFEDEGCKEYIYKPKLTPLEISQRLKLKYSKFGSENVKMIQDSKVPKDLDSKY 1863  
 Db 2300 NDVIDIND-----INDTNDINQNHQNSSDK-----KDSKRTANTGNTIRY 2340

Qy 1864 AYIQVTHVFPDEKELOERK---TEPERS 1890  
 Db 2341 AGAIFGLVTFVIRIRIEKDKDEMFDKS 2371  
 RESULT 13  
 A27605  
 dystrophin, muscle - human  
 N.Alternate names: Duchenne muscular dystrophy protein  
 C.Species: Homo sapiens (man)  
 C.Date: 19-Nov-1988 #sequence\_revision 27-Jun-1994 #text\_change 16-Jun-2000  
 C.Accession: A27605; S07710; A27162; S05291; A40134; S06051; S10346; S02242; S02  
 R.Koenig, M.; Monaco, A.P.; Kunkel, L.M.  
 Cell 53, 219-228, 1988  
 A.Title: The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein.  
 A.Reference number: A27605; MUID:88194521; PMID:3282674  
 A.Accession: A27605  
 A.Molecule type: mRNA  
 A.Residues: 1-3685 <KOE>  
 A.Cross-references: GB:M18533; NID:G181856; PIDN:AA53189.1; PID:G181857  
 R.Rosenthal, A.; Speer, A.; Billwicz, H.; Cross, G.S.; Forrest, S.M.; Davies, K.E.  
 Nucleic Acids Res. 17, 5391, 1989  
 A.Title: Two human cDNA molecules coding for the Duchenne muscular dystrophy (DMD) locus  
 A.Reference number: S07710; MUID:89345106; PMID:2668885  
 A.Accession: S07710  
 A.Status: nucleic acid sequence not shown; translation not shown  
 A.Molecule type: mRNA  
 A.Residues: 1-132, 'P', 134-622, 'I', 624-783, 'G', 785-1196, 'P', 1198-1376, 'N', 1378-1468, 'Q', 1  
 A.Cross-references: EMBL:X14298; NID:G30845; PIDN:CAA32479.1; PID:G30846  
 A.Note: This sequence was submitted to the EMBL Data Library, February 1989  
 R.Koenig, M.; Hoffman, E.P.; Bertelson, C.J.; Monaco, A.P.; Feener, C.; Kunkel, L.M.  
 Cell 50, 509-517, 1987  
 A.Title: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary  
 A.Reference number: A50897; MUID:87273512; PMID:3607877  
 A.Accession: A27162  
 A.Molecule type: mRNA  
 A.Residues: 1-497 <KOE>  
 A.Cross-references: GB:M18533  
 R.Cross, G.S.; Speer, A.; Rosenthal, A.; Forrest, S.M.; Smith, T.J.; Edwards, Y.; Flint,  
 EMBO J. 6, 3277-3283, 1987  
 A.Title: Deletions of fetal and adult muscle cDNA in Duchenne and Becker muscular dystro  
 A.Reference number: S01263; MUID:8811512; PMID:3428261  
 A.Accession: S05291  
 A.Molecule type: mRNA  
 A.Residues: 404-556, 'V', 558-610, 'K', 612-622, 'I', 624-664, 'W', 665-783, 'G', 785-1137, 'P', 'N' <C  
 A.Cross-references: EMBL:X06178  
 A.Note: 475-11e and 529-Glu were also found  
 R.Hoffman, E.P.; Monaco, A.P.; Feener, C.C.; Kunkel, L.M.  
 Science 238, 347-350, 1987  
 A.Title: Conservation of the Duchenne muscular dystrophy gene in mice and humans.  
 A.Reference number: A40134; MUID:88018015; PMID:3659917  
 A.Accession: A40134  
 A.Molecule type: mRNA  
 A.Residues: 491-1207 <HOF>  
 A.Cross-references: GB:M18533  
 R.Blonden, L.A.J.; den Dunnen, J.T.; van Paassen, H.M.B.; Wapenaar, M.C.; Grootscholten,  
 Nucleic Acids Res. 17, 5611-5621, 1989  
 A.Title: High resolution deletion breakpoint mapping in the DMD gene by whole cosmid hyb  
 A.Reference number: S06051; MUID:89345155; PMID:2569720  
 A.Accession: S06051  
 A.Status: translation not shown  
 A.Molecule type: DNA  
 A.Residues: 2147-2204 <BLO>  
 A.Cross-references: EMBL:X15495; NID:G30829; PIDN:CAA3518.1; PID:G1335049  
 R.Speer, A.; Billwicz, H.; Hutch, A.; Couelle, C.; England, S.; Love, D.; Davies, K.E.  
 submitted to the EMBL Data Library, February 1990  
 A.Reference number: S10346  
 A.Accession: S10346  
 A.Molecule type: DNA  
 A.Residues: 2438-2480 <SPE>  
 A.Cross-references: EMBL:X51934  
 R.Chamberlain, J.S.; Gibbs, R.A.; Ranier, J.E.; Nguyen, P.N.; Caskey, C.T.

Nucleic Acids Res. 16, 11141-11156, 1988  
A>Title: Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA  
A'Reference number: S02109; MUID:89083552; PMID:3205741  
A'Accession: S02243  
A'Status: preliminary; translation not shown  
A'Molecule type: DNA  
A'ResIdues: 665-722 <CH3>  
A'CROSS-references: EMBL:X13045; NID:g30825; PIDN:CAA31451.1; PID:g1335048  
A'Accession: S02242  
A'Status: preliminary; translation not shown  
A'Molecule type: DNA  
A'ResIdues: 2098-2146 <CH2>  
A'CROSS-references: EMBL:X13046; NID:g30827; PIDN:CAA31452.1; PID:g809549  
A'Accession: S02244  
A'Status: preliminary; translation not shown  
A'Molecule type: DNA  
A'ResIdues: 2147-2204 <CH3>  
A'CROSS-references: EMBL:X13048; NID:g30833; PIDN:CAA31454.1; PID:g1335051  
A'Accession: S02109  
A'Status: preliminary; translation not shown  
A'Molecule type: DNA  
A'ResIdues: 2305-2365 'K' <CH4>  
A'CROSS-references: EMBL:X13047; NID:g30831; PIDN:CAA31453.1; PID:g1335050  
A'Glnjaat, I.H.B.; van Paasessen, M.H.M.B.; den Dunnen, J.J.T.; van Ommen, G.G.J.B.  
submitted to the EMBL Data Library, March 1992  
A'Description: Sequence of Duchenne muscular dystrophy gene exon 60, located directly 5'  
A'Reference number: S23736  
A'Accession: S23736  
A'Status: preliminary  
A'Molecule type: DNA  
A'ResIdues: 2980-2995 'K', 2997-3028 <GIN>  
A'CROSS-references: EMBL:X11860  
R.Hellig, R.; Lemaire, C.; Mandel, J.L.  
Nucleic Acids Res. 15, 9129-9142, 1987  
A'Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection of a cor  
A'Reference number: S09071; MUID:88067745; PMID:2825128  
A'Accession: S09071  
A'Molecule type: DNA  
A'ResIdues: 'SGGSHWTHGSLRYLPLTLI', 218-277 <HEI>  
A'CROSS-references: EMBL:X06293; EMBL:Y00494  
A>Note: Sequence N-terminal of residue 218 correspond to a putative exon  
R.Roberts, R.G.; Coffey, A.J.; Bobrow, M.; Bentley, D.R.  
Genomics 16, 536-538, 1993  
A'Title: Exon structure of the human dystrophin gene.  
A'Reference number: I54186; MUID:93300536; PMID:8314593  
A'Accession: I54186  
A'Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A'Molecule type: DNA  
A'ResIdues: 984-1411 <RE3>  
A'CROSS-references: GB:L05642; NID:g181892; PIDN:AA474506.1; PID:g950344  
A'Accession: I68509  
A'Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A'Molecule type: DNA  
A'ResIdues: 1776-1913 <RE2>  
A'CROSS-references: GB:L05646; NID:g181896; PIDN:AA474507.1; PID:g950345  
A'Accession: I68510  
A'Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A'Molecule type: DNA  
A'ResIdues: 2850-2979 <ROB>  
A'CROSS-references: GB:L05649; NID:g181899; PIDN:AA474508.1; PID:g950346  
R.Roberts, R.G.; Coffey, A.J.; Bobrow, M.; Bentley, D.R.  
Genomics 13, 942-950, 1992  
A'Title: Determination of the exon structure of the distal portion of the dystrophin gen  
A'Reference number: I54175; MUID:92372062; PMID:1505385  
A'Accession: I54175  
A'Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A'Molecule type: DNA  
A'ResIdues: 2980-3685 <RES>  
A'CROSS-references: GB:M66903; NID:g181881; PIDN:AA43779.1; PID:g457519  
R.Ehnenpfele, J.; Hillner, M.; Junkes, B.; Pfordt, M.; Schwinger, E.; Voeberg, H.P.  
Genomics 10, 551-557, 1991  
A'Title: Analyses of a dystrophin gene deletion by amplification of mRNA isolated from D  
A'Reference number: I54166; MUID:91365360; PMID:1889805

A'Accession: I54166  
A'Status: translated from GB/EMBL/DBJ  
A'Molecule type: mRNA  
A'ResIdues: 2250-2254 <RE4>  
A'CROSS-references: GB:S54699; NID:g235303; PIDN:AA19754.1; PID:g235304  
R.Feeney, C.A.; Koenig, M.; Kunkel, L.M.  
Nature 338, 509-511, 1989  
A'Title: Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy  
A'Reference number: S03902; MUID:89181947; PMID:2648158  
A'Accession: S03902  
A'Molecule type: mRNA  
A'ResIdues: 'MED', 12-32, 3377-3408 <FE8>  
A'CROSS-references: EMBL:X15148  
C'Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl  
C'Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro  
C'Genetics:  
A'Gene: GDB:DMD  
A'CROSS-references: GDB:119850; OMIM:310200  
A'Map position: Xp21.2-Xp21.2  
A'Introns: 11/1, 31/3, 62/3, 88/3, 119/3, 177/2, 217/1, 277/3, 320/3, 383/3, 444/2, 494/  
3, 3055/1, 3075/2, 3096/1, 3121/1, 3188/2, 3217/1, 3269/3, 3325/2, 3362/3, 3408/2, 3421/  
A>Note: the list of introns is incomplete  
C'Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystrop  
C'Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; leuci  
elix  
F.154-233/Domain: alpha-actinin actin-binding domain homology <ACT>  
F.253-327/Region: hinge  
F.338-447/Domain: spectrin/dystrophin repeat homology <SP01>  
F.448-556/Domain: spectrin/dystrophin repeat homology <SP02>  
F.558-667/Domain: spectrin/dystrophin repeat homology <SP03>  
F.668-717/Region: hinge  
F.718-828/Domain: spectrin/dystrophin repeat homology <SP04>  
F.838-934/Domain: spectrin/dystrophin repeat homology <SP05>  
F.938-1045/Domain: spectrin/dystrophin repeat homology <SP06>  
F.1047-1154/Domain: spectrin/dystrophin repeat homology <SP07>  
F.1156-1263/Domain: spectrin/dystrophin repeat homology <SP08>  
F.1265-1367/Domain: spectrin/dystrophin repeat homology <SP09>  
F.1372-1477/Domain: spectrin/dystrophin repeat homology <SP10>  
F.1478-1568/Domain: spectrin/dystrophin repeat homology <SP11>  
F.1570-1676/Domain: spectrin/dystrophin repeat homology <SP12>  
F.1678-1782/Domain: spectrin/dystrophin repeat homology <SP13>  
F.1784-1875/Domain: spectrin/dystrophin repeat homology <SP14>  
F.1876-1982/Domain: spectrin/dystrophin repeat homology <SP15>  
F.1984-2101/Domain: spectrin/dystrophin repeat homology <SP16>  
F.2103-2208/Domain: spectrin/dystrophin repeat homology <SP17>  
F.2210-2316/Domain: spectrin/dystrophin repeat homology <SP18>  
F.2327-2423/Domain: spectrin/dystrophin repeat homology <SP19>  
F.2424-2470/Region: hinge  
F.2471-2577/Domain: spectrin/dystrophin repeat homology <SP20>  
F.2579-2686/Domain: spectrin/dystrophin repeat homology <SP21>  
F.2688-2802/Domain: spectrin/dystrophin repeat homology <SP22>  
F.2804-2931/Domain: spectrin/dystrophin repeat homology <SP23>  
F.2933-3040/Domain: spectrin/dystrophin repeat homology <SP24>  
F.3041-3112/Region: hinge  
F.3055-3092/Domain: WW repeat homology <WW1>  
F.3080-3360/Region: cysteine-rich  
F.3506-3527/Region: leucine zipper motif  
F.3572-3593/Region: leucine zipper motif  
Query Match 2.0%; Score 213.5; DB 1; Length 3685;  
Best Local Similarity 17.1%; Pred. No. 0.0034;  
Matches 404; Conservative 347; Mismatches 830; Indels 777; Gaps 102;  
QY 12 ETRKTRALSKPGTAIELRQSVSEVRGVSYLAKPKLIPLDYENAVYQKQTQIINDCLR 71  
DB 326 EDKFSGLSMEVNLRYQTALEVLISLSDTL-----QAQGRISNDV-- 372  
QY 72 EML---LFPYDQQTALIRQGRY-----ICSTYPAKAREAGSLPTECKTKYNS 119  
DB 373 EVVNDQGFTHGMYMDLPAHQGRVGNITLQASKLIGTKSEDETR---VQEQNNILNS 429  
QY 120 DMHIVNY---KYEDSGEPRQLPNKVKLKLPHVHYEVDERVKDED---AASLGSQK 172



Db 430 RRECLRVASMEKOSNLHRLVMDLQNO--KLKELNDMLTKTEERTKREKEEPLGPDLEDUK 487  
 Qy 173 GGIITKGMVLYKGMNNAISVTRMSFRRRPHLLQLODGSYKPEFLKDLOKEPKGSIFLGF 232  
 Db 488 ROVOQKHVLOEDLEQOVRNSILT-----HMVVVBESSGDHATLAEOLK----- 534  
 Qy 233 LYGVSRNNKVRFAFELKMQDKSYLLAADSEVEEMWITILINKL-----QUNPE 284  
 Db 535 VLGDRAN--ICWT-----EDRWVLLODILKKORLTHEBCULFS 572  
 Qy 285 AAMQERANDSHEDDEQSLKESG--SGLDLYPELAKSA--REAEIKLKEBSRYKLFYLDP 342  
 Db 573 AMLSEK-----EDAVNNKHITTFKDOHEMLSSLOKLAVLKADLEKKQOSMKYSLXQ 625  
 Qy 343 D-AQXLDPSABREVSFEKEFKRLIVNC--NDLSNLOCCVANESEPTNVEPFVYL 400  
 Db 626 DLSTLKNKSVTOKTEAMLDNF-----ARCMDLVOKLEKSTRAQISOAVTT--TQPSLTOT 679  
 Qy 401 SLFD-----IKY-----NRKISADPHV-----DLNHSVROMIATYS 432  
 Db 680 TWMEVTVTTTRQOLVKHAQBELPRPPQKKQOIVDSEIRKRLVDITLHSMI----- 735  
 Qy 433 PALMNGSGEPTGALGILHEAMQYPKQGISVTCRPHDI--FLVARIEKVLQGSITTHCA 491  
 Db 736 -----TRG-----EAVLOGBEPALFRKEGNFSDLKEKVAIER-----EKA 771  
 Qy 492 EPMKSSDSSKVAVOKYLK-----NAKQACORLQGYRMPFPAARATTLFMDASGND 541  
 Db 772 EKFRKLODSRSRQALVEQVNEGVNADSTKQASELNSRWIEFC-----QLSERLNMIE 827  
 Qy 542 KNARFSAIYRQ-----DSNKLSDMDMLKLLADFRKPRKAKLRYILGNLDITIDNYS--SD 595  
 Db 828 YQNNIIAFYNOLOQLQEMOTTLAENWLKI-----QPTPSEPTAKSQKLCIDEVVRISLG 882  
 Qy 596 FRPVYNSVITPTKQFETCSKTPITTEVEEVPFCIRPHQTPYITTYNMLVYPRYKLYDSQ 655  
 Db 883 LQPIERLKIQSIALKEKQSGPFLDAD-----FVAFTNHF-----KQVPSDQ 926  
 Qy 656 KSPAKARNIAICIEFDSDEBDSQPLKCTYGRGQVFTPSAFAALHNNHONDEPDEIX 715  
 Db 927 ABEKELOTF-----DTLP-----DMRYOETWSAIRTVQO-----SEK 961  
 Qy 716 IELPTOLHEKHNLLTFEPVSCDSSKSGTKKRDVETQVGSWLPRLKDGVRVSECHI 775  
 Db 962 LSP-----QLSVTDYELMEORLDELALQSSLOQSGLYL-----STYK 1004  
 Qy 776 PVSANLPSGYLQYELGMRHNGPELKWVDSGKPLKISTHLYSTVYTTODOLHNFQYC 835  
 Db 1005 EMSKAPR-----EISRKYQSEFEIEG--RMKCLSSOLV-----EHC 1040  
 Qy 836 QCTESGAOLGNELVYKLSLHMEGHVMIAPLPTLNQLPRLTRATQEBVAVNVTRYI 895  
 Db 1041 QKLE-----EQMKLKRKIQNHIO-----TLKMMAEVDVYLKEBMPALGDSLTL 1084  
 Qy 896 IHVVAOCH-----EBGLSEHLSUYKYAYKAPRYVASEYKT----- 931  
 Db 1085 KQOLQKCRLLVSDIOTIOPSLSVNGGQKIKNEAPREFASRLJETELKELNTQMOMHMQO 1144  
 Qy 932 -----VHEBLTKSWTTILKPSADFLTSNKLKTSWFFDYV 967  
 Db 1145 VYARKKALGKLEKTVSLQKLDSEMHMWTQAEELYERDFEYKTPDELOK-----AVEEM 1200  
 Qy 968 IKSMACHLLENSKVLLRNQRPASYNHAVETVNMMLPHIOTQKPDNPA--SKAN----- 1023  
 Db 1201 KKAKEBAQOKKAKVLLT-----ESVNSVIAQAPVAQBALKEELTETLTNYQMLC 1251  
 Qy 1024 -----HSLAVFIKRCFTFMDR--GFVFOINNYISCFAPGDKTLFEYK 1065  
 Db 1252 TRLNGKCTLBEVWACMHELSTYLBKANKMLNEVEFKLTTEH-----IPGABEISEV-- 1305  
 Qy 1066 FEFLLRVVCHNEHTIPLNLPMPFGKRIQRYODIQLDYSLTDFECRNHPLVGLLREVGTA 1125  
 Db 1306 LPSLENTMBHSEBNDP-----NOIRILAQTLTDGGVWDE-----LINELETF 1347

Qy 1126 LOEFREVLIAISVLKNNLLKHSFDDRYASRSHQARIATVILPLFGLLIENVRINVDV 1185  
 Db 1348 NSRMELIEBAVRK-----LEQSIQSQOETEK 1377  
 Qy 1166 SPPPVNAGMTYDESLA---LPAVNPVLTPOKSGTLDNSLHKDLGALSGIASPYTSTP 1242  
 Db 1378 SLHLIQESLTFIDKOLAIVADKYDAQMPQDAQIQGDL----- 1417  
 Qy 1243 NINSVRNADSRKSLISTSGNSLPRNSEKNSLDKQOOSTLGNVYRCOKLDQSEIKS 1302  
 Db 1418 -----TSHBISLEB-----MKKHQKGAARVLSQIDVAKQLQD 1453  
 Qy 1303 LLMCFYILKMSWDDALFTYNNKASTSELMDFITSEVCLHQFOYMGARYIARTGMHAR 1362  
 Db 1454 VSM-----KFRLOKAPNFE-----RKQESKMLDVEKYNL 1486  
 Qy 1363 LOOLGSLDNLST---FNHSYGHSDADVHQSLEANIATEVCLTALDLSLFTLAFKNQ 1419  
 Db 1487 ALEKTSVEQEVVQSQLNHCN-----LYKSLSE--VKSEV-----EM 1521  
 Qy 1420 LADHGHNPLMKVPVYLCFLOKHSE-----TLKXVFTALRSLYKPESTFEG 1470  
 Db 1522 VIKTRQ-----IVQKQTEPNKELDERVTALKLHYELGAKVTERQO----- 1564  
 Qy 1471 RADMCALCYELIKCSNGLS--SIRTEASOLLYFLMRNPFYTGKSFV----- 1518  
 Db 1565 -----OLEKL--KLRSKRKKNVNLTEWLAADIMELT--KRSNABECPNSLJDEV 1611  
 Qy 1519 -----RTHLOVITISVQSLADVVGIGTRFOQSLSTINN-----CANSRLI 1560  
 Db 1612 AMGRATQKEIEKQKHNLSITEVEGALKTVGLKKEVLEDKSLINSNMIAVTSABEML 1671  
 Qy 1561 -----KH--TSFSSDVVDLTKRI--RTVLMAIQMKEHNDPEMLVDLOQSLKSVAST 1610  
 Db 1672 NLLLEYQKMETFPDQNDVHTIKWIIQADTLDESEKKRQOQEDVLKTKLKLNDIRPV 1731  
 Qy 1611 PELARKTMDSMARIHVNKGDLSBAMCYVHTALVABYLTRKAAQVEMPP--LPHSHA 1668  
 Db 1732 DSTR-----DOANLMAANGD-----HCKRLV-----EPQISELNHRFAA 1766  
 Qy 1669 CLRSRGVFRQCTAFRVITPNIDE-----EASMEVDVGMQDVAFNEV-----LME 1716  
 Db 1767 ISHRIKTGKASIPRLKELQFNSDIQKLEPRLAEIQGVNLKEEDFNKDMNEDNGTYKE 1826  
 Qy 1717 LLEQCADLGMKAEKRELADIYKLIIPYEXRROBERLAHYDTHRAY-----SKV 1768  
 Db 1827 LLQR--GDNL-----QQRITD-----ERRREBEIKIKOULLQTHGNLKDLSQRKKA 1872  
 Qy 1769 TEVMSGRLLGTVRVAFQQAQYOPTDSETDVEGFEDDEGKEYIYKEPK--LTPUSE 1827  
 Db 1873 LEISHQ-----WYQYKRAODLLKLDLIDIEKTLAALPREPRDRKIKE 1914  
 Qy 1828 ISORLLK---LYSDKFGSENVKMIODSGKVP-----KOLDSRYAYIQTHTVIP 1874  
 Db 1915 IDRELQKKKEELNAAVRQAEGLSEGDAMAAYEPQIOLSKEMREIESKFA----- 1964  
 Qy 1875 FDEKLOERKTEPFSHNIRRFMEBMPPTQVKGKQGVSEBQCKRTILTALHCPRYVKKR 1934  
 Db 1965 -----QERRLNFAOIHTVR-----BETMMVMTEDMPELISY 1995  
 Qy 1935 IPVMTQHTNDPFEVALIDEMSKVAELRQCSSAEVDMIKLOLQLOG--SVSVQVNAQ-- 1991  
 Db 2053 DIHSHKTAALQSATPVRVRYKLOEALSQLDQWEMKVNMGYKXQCFRFSVEKMRFRFYD 2112  
 Qy 2024 VBAQQAALANVERLIKQOLEYQOEBMKANYEMAKELSE-----IMHEQICP 2070  
 Db 2113 IKIFQWMTAEBOFLRKTQIPEMNE--HAKYKMYLKEODGIGQRTVVRITNAGEBIIQ 2171



QY 2071 LEKKT--SVLPNSLHIEN 2086  
 Db 2172 OSSKTDASIIQEKGLSIN 2189

## RESULT 14

114156  
 kinesin-related protein - African clawed frog  
 C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #ext\_change 11-May-2000  
 C/Accession: T14156  
 R/Mood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
 Cell 91, 357-366, 1997

A/Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosc  
 A/Reference number: Z17893; NUID:98028574; PMID:9363944  
 A/Accession: T14156  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Molecules: 1-2954 <WOO>  
 A/Cross-References: EMBL:AF027728; NID:92586070; PID:92586071; PIDN:AAC60300.1  
 C/Genetics:  
 A:Gene: XENP-E  
 C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 1.9%; Score 212.5; DB 2; Length 2954;  
 Best Local Similarity 18.2%; Pred. No. 0.0027;  
 Matches 424; Conservative 336; Mismatches 724; Indels 851; Gaps 111;

QY 32 SVSEVVRGSLVLAKEPLDPEYIVQKTIQLNDCLREMLPFYDPTALLRQGR 91  
 Db 335 STAKKIVRNT-----PQVNEVLDEALALKYRKELID-----LKQLE 361  
 QY 92 YICSTVPAKA-----EEBAQSLFVTECIKTNSD--WHLVNYKYEDYSGEPRQ----- 137  
 Db 362 NIESSSETQAQMAKEBHQAELKQHLKEBEDRIMHLTNIVAS--SGESQDDQVYRK 420  
 QY 138 ----LPNKVRK-----LDKLPVHVY-----EVDEEVDKD---ED 164  
 Db 421 RRVTAAPGKIQNSLHASGVDFDMLSRILPGNFSKAKAFSDMPSFEIDSVCTESDFD 480  
 QY 165 AASLSQKCGITKHCMLYGNMNSAISYTMBSFKRRFHLIQLGDSYFEBLKQLQKRP 224  
 Db 481 ALSMDSNDC-----IDAEWNLASKVTHREKTSIHQSMIDFGQISDSVQF--HDSKSN 531  
 QY 225 KCSIFLGFIYGVSPFNKRYRFAFELKMDKSSYLLADSE-----VEMEEMI 272  
 Db 532 Q-----LQYLPKDSGMAEGRKASFKEKITSLOQLOQSKSEKELVQSEELKIALDEBL 587  
 QY 273 TI-----LNKILQLNFEAMQEKRN---GESHEDDEQSKLEGS---GSGI 311  
 Db 588 SVKAKNLEWNTNSRHSINAVQOTDVEKEVVRKEMSVLGLSGYNASNDLQDSSVDGRKL 647  
 QY 312 DSYLPELAKSAREAEIKLKSSESRVYKLFYDPPDAQKIDFSSAEPEVYKFEFKGKRLV-- 369  
 Db 648 SSSHDECEIHRKMLBOKIVD-----LEEFTEINMKKSEENDKQSSQEDPMEISIQICE 699  
 QY 370 -----KCNLDL-----FYLQCCVAENERGPTTNVEPFYVTLSPRIKXNRKISADPHND 418  
 Db 700 AIMAEKANLBEELALMRDNPNIILENET-----LKREIADLESLEKNO---E 745  
 QY 419 LNHFSVRQMIATTSPALMNGSGPETQSALRGILHEAMQYPRQGIIFSVCPPHPIFLVAR 478  
 Db 746 TNEFELLE-----KETQKE-----HEAQI----- 764  
 QY 479 IEKVLQGSITTHCAEPYMKSSDSKVAQKVLKNAKQAQRIQGYRMPFAMAATLFRDASG 538  
 Db 765 IHEI--GSI-----KKIVENAEVNTQ----- 783  
 QY 539 NLDKARFSAIYRQDSNKLSDNDMLKLADPRKPEGMALPVILGNDITIDNVSDFPN 598  
 Db 784 NLEBDL-----ETTKLLKEOEIO-LAELEK--RADNLOKKVRNFDLSVMGDSB--- 830  
 QY 599 YNNSYIPIKQPEFETSKPTITEVEBEFVDCIPGHQPYTIYTNHLVYVYKYLKYSQK-- 656

Db 831 -----KICEB---IFOLKQSLSDAEAYTR-----DAQKEC 857  
 QY 657 SPKARNIAIICIFPDSDDEDSQPLKCIYGRPGGVFRSAFAVLHNNHNPBFDEIKI 716  
 Db 858 SPLSEN---LELKEKEDTS-----NMNTNQ--K 881  
 QY 717 ELPTQLHEKHHLILTFPFAVSCDNSKSTKRDVETQGVSWPLRLXQGRVVTSEQH1P 776  
 Db 882 EKASLSPFKQL-----ETEKSYKKQKGA-----D 905  
 QY 777 VSANLPSGYLGVOELGMRHYGPEIKWVGGKPL--KISTHLVSTYTTTQDQHLNFPQY 834  
 Db 906 LQKEIQSAF-----NEINYNG--DLAGVPRDLISR-- 936  
 QY 835 CQKTESGAQALGNELVYKLSLAMEGVM-----IAPPLTLNQDFRLVTAQCE----- 885  
 Db 937 --ELEKVSSEFSKOLEKALBEKNALNENVTCLSEYKFLPNEVECKNQISKASEIMILK 994  
 QY 886 ---EVAVANVTVIIHVVAQCHEEGLE-----SHLSRYKVAVYKAPYVASEYKTVHEBLT 937  
 Db 995 QEGHSASIIKQELIMQOSEQILQLDEVTHQSKVQOT--BEQYL--EMKGNHDLF 1050  
 QY 938 KSMITILKPSADFLTSTNKLKYSWFFDVLIKMAQHLIE--NSKVLLRNQRPASVYH 995  
 Db 1051 EKYIRNKSEADLIREMENLKGTMESVEKIADTYGHELEETIRDKELHKKY---PFQ 1107  
 QY 996 AVETVV-----NMIMP--HIQKPRNDPEAKNANHSLAVIKICFTMDGFPVKQINN 1048  
 Db 1108 AMQITFPIPTPLSDSLPPSKLVEGNSQDPIEINDYNNLTALATER-----NN 1153  
 QY 1049 YISCFAPDPTLTFEYKFEFLRVVNCNHHYIPLNLP-----PFGKRIOR--YQDL-- 1098  
 Db 1154 IMWC-----LETENSLK-----EQVIDLNTQLOSLQAQSIKSDLDKQKPDLEB 1199  
 QY 1099 QLDVSLDPEFCRNHFL-----VGLLREYGTALQEFREYRLAISVLKNLIXH-- 1147  
 Db 1200 EVKLLLEMEELKGLHLDQSLSIEKLOLENTLEVTLEQLTQCE-----EMKQITIERNEL 1252  
 QY 1148 --SPDRYASRSHQARITVLPFLGILLIENVQRIINRVADVPPFNAGMTYKDSLAPA 1205  
 Db 1253 QTNFEDLAEHD-----SLKQDISENIEQ-----SLETQDE--IRA 1286  
 QY 1206 VNPVLPQKSGTLDNLSLHKDLGAISGYASTTSTP-----INSVRNADRG--- 1254  
 Db 1287 AQEHLREQK--QLVDSFQQLDSCVGISSEPHDAVANQEKVSLGEVNSLSQSEMLRGERD 1344  
 QY 1255 -----SLIS-----TDSGNSLPERNSEKNSLDRHQOOSTLGNSVVRCDKLDQ 1297  
 Db 1345 ELQTSKALVELELLRAHVKSVEGENL--EITKLNGLER--BILGKS-----EE 1391  
 QY 1298 SEIKSLMCFIYILKSMDDALPTYMANKASISELMDFTTISEVCLHQCQYNGKRYIARTG 1357  
 Db 1392 SE-----VKSMLDEN-----KEDNNKLUKSE----- 1411  
 QY 1358 MHAHALQOLGSLNLSLTFNHSYGHS-----DADVLHQSLEANLATEVC-----LTALD 1406  
 Db 1412 -----QAEYSKSENOFSLEEVFGSQKLVDEIVLKAQKAAEBLEIKDDYFLVQTA 1467  
 QY 1407 TLSLFTLAFKNQLADHGHPMLKVPDVYLCPLOKQSEFALKNVFTALBSLYKFPST 1466  
 Db 1468 NTNILVEGKLETPLOQDHBEDSIDRSEMEI---KVLGKLENNQVYLLER----- 1514  
 QY 1467 FYEGADNCALCYEILK-----CNSKLSISRTASQULYFLMKNPNPYTGK- 1514  
 Db 1515 LOEBKELGKSNKL--EIIQKEMETSVLKLDDLOOKLESLSB---NIIKEMENDITTKH 1567  
 QY 1515 KSFVTRHLQVILISVSQLADVAVIGTFRFOOSLSINN--CANSDR--LIKHTSPSSD-V 1569  
 Db 1568 HSDTQAOQ-----KTQEBLOLAKRLAIADNDCPTIOBEKESALCV 1609  
 QY 1570 KDLTRIRITVLMATQAOMKEHNDPEMLV---DLQYSLAKSYASTPELAKTMDLSMA--- 1622



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:02:46 ; Search time 21.9153 seconds

(without alignments)  
3987.653 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936

Sequence: 1 MSQPLPLPASAEKTRKTRAL.....ISGTPSTVHGHTSSSSSV 2107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213.5	2.0	3685	1 DMD_HUMAN	P11332 homo sapien
2	203	1.9	2077	1 TEGU_HSV6G	P30002 human herpes
3	199	1.8	3210	1 CENF_HUMAN	P49454 homo sapien
4	196.5	1.8	2748	1 NDM1_YEAST	Q00402 saccharomyc
5	195	1.8	2077	1 TEGU_HSV6U	P52340 human herpes
6	193	1.8	1225	1 G160_MOUSE	P55837 mus musculu
7	190.5	1.7	4568	1 DYHC_CAEEL	Q19020 caenorhabdi
8	187.5	1.7	2869	1 RBP1_PLAVB	Q00798 plasmodium
9	186	1.7	1818	1 HMM2_MYCPN	P75471 mycoplasma
10	185	1.7	2663	1 CENB_HUMAN	Q02324 homo sapien
11	185	1.7	1557	1 DVAL1_DICVT	Q24702 dictyocaulu
12	184	1.7	4385	1 YP73_CAEEL	Q09222 caenorhabdi
13	183.5	1.7	1901	1 AKI1_HUMAN	Q90444 homo sapien
14	183	1.7	5327	1 ACF7_MOUSE	Q94920 mus musculu
15	181.5	1.7	1957	1 YD86_SCHPO	Q10411 echizosacch
16	181	1.7	2017	1 YU77_YEAST	P39526 saccharomyc
17	181	1.7	3680	1 DMD_CANFA	Q97592 canis famli
18	179.5	1.6	1966	1 MYSE_CAEEL	P02566 caenorhabdi
19	178.5	1.6	1679	1 Y109_YEAST	P40457 saccharomyc
20	176.5	1.6	2273	1 HPA1_YEAST	P32874 saccharomyc
21	175	1.6	1805	1 HMM2_MYCGE	P47460 mycoplasma
22	172	1.6	3056	1 ATM_HUMAN	Q13315 homo sapien
23	171.5	1.6	3678	1 DMD_MOUSE	P11531 mus musculu
24	171	1.6	1151	1 XPO4_MOUSE	Q96910 mus musculu
25	169	1.5	1526	1 MYB2_SCHPO	Q90416 gallusgallu
26	168	1.5	3660	1 DMD_CHICK	P11533 gallusgallu
27	167.5	1.5	1875	1 MLPI_YEAST	Q02455 saccharomyc
28	167	1.5	1151	1 XPO4_HUMAN	Q90452 homo sapien
29	166	1.5	2136	1 YCF2_MARPO	P09975 marchantia
30	165.5	1.5	2704	1 BPA1_HUMAN	Q03001 homo sapien
31	165	1.5	2469	1 TEGU_HSV5A	Q01056 herpesvirus
32	164	1.5	2059	1 TEGU_HSV7J	P52362 human herpes
33	164	1.5	3911	1 AKA9_HUMAN	Q99996 h-a-kinase

34	164	1.5	4092	1 DYHC_YEAST	P36022 saccharomyc
35	164	1.5	4128	1 PRKD_MOUSE	P97313 mus musculu
36	162.5	1.5	2875	1 RRP1_TSVW1	P28976 tomato spot
37	161.5	1.5	5430	1 ACF7_HUMAN	Q94920 mus musculu
38	160	1.5	1928	1 MYSL_YEAST	P03964 saccharomyc
39	159	1.5	2026	1 CYAA_YEAST	P03678 saccharomyc
40	159	1.5	2472	1 SPCN_RAT	P16086 rattus norv
41	158	1.4	4563	1 APB_HUMAN	P04114 homo sapien
42	157	1.4	1420	1 YMBE_YEAST	Q03496 saccharomyc
43	157	1.4	2144	1 BP28_HUMAN	Q95883 homo sapien
44	157	1.4	2349	1 TPR_HUMAN	P12270 homo sapien
45	157	1.4	3080	1 PDG_ZYMYC	P18479 z genome po

## ALIGNMENTS

RESULT 1	DMD_HUMAN	STANDARD;	PRT; 3685 AA.
ID	DMD_HUMAN		
AC	P11532; Q14169; Q14170;		
DT	01-OCT-1989 (Rel. 12, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DT	15-0UN-2002 (Rel. 41, Last annotation update)		
DB	Dystrophin.		
GN	DMD.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OK	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88194521; PubMed=3282674;		
RT	Koenig M., Monaco A.P., Kunkel L.M.;		
RT	"The complete sequence of dystrophin predicts a rod-shaped		
RT	cytoskeletal protein.";		
RT	Cell 53:219-228(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89345106; PubMed=2668885;		
RA	Rosenthal A., Speer A., Billowitz H., Cross G.S., Forrest S.N.,		
RA	Davies K.E.;		
RT	"Two human cDNA molecules coding for the Duchenne muscular dystrophy		
RT	(DMD) locus are highly homologous.";		
RT	Nucleic Acids Res. 17:5391-5391(1989).		
RN	[3]		
RP	SEQUENCE OF 1-497 FROM N.A.		
RX	MEDLINE=87273512; PubMed=3607877;		
RA	Koenig M., Hoffman B.P., Bertelson C.J., Monaco A.P., Feener C.,		
RA	Kunkel L.M.;		
RT	"Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and		
RT	preliminary genomic organization of the DMD gene in normal and		
RT	affected individuals.";		
RT	Cell 50:509-517(1987).		
RN	[4]		
RP	SEQUENCE OF 404-1137 FROM N.A.		
RX	MEDLINE=8811512; PubMed=3428261;		
RA	Cross G.S., Speer A., Rosenthal A., Forrest S.M., Smith T.J.,		
RA	Edwards Y., Flint T., Hill D., Davies K.E.;		
RT	"Deletions of fetal and adult muscle cDNA in Duchenne and Becker		
RT	muscular dystrophy patients.";		
RT	EMBO J. 6:3277-3283(1987).		
RN	[5]		
RP	SEQUENCE OF 665-722; 2098-2204 AND 2305-2366 FROM N.A.		
RX	MEDLINE=89083552; PubMed=3205741;		
RA	Chamberlain J.S., Gibbs R.A., Ranier J.A., Nguyen P.N.,		
RA	Caeskey C.T.;		
RT	"Deletion screening of the Duchenne muscular dystrophy locus via		
RT	multiple DNA amplification.";		
RT	Nucleic Acids Res. 16:11141-11156(1988).		
RN	[6]		
RP	SEQUENCE OF 2147-2204 FROM N.A.		
RX	MEDLINE=89345155; PubMed=2569720;		

RA Blonden L.A.J., den Dunnen J.T., van Paaseen H.M.B.,  
 RA Wapenaar M.C., Grootecholten P.M., Ginjaar H.B., Bakker E.,  
 RA Pearson P.L., van Ommen G.J.B.;  
 RT "High resolution deletion breakpoint mapping in the DMD gene by whole  
 RT cosmid hybridization.";  
 RT Nucleic Acids Res. 17:5611-5621(1989).  
 RL (7)  
 RP SEQUENCE OF 2305-2364 FROM N.A.  
 RA Hutch A., Will K., Speer A., Bauer D.;  
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
 RN (8)  
 RP REVIEW ON DMD POINT MUTATION VARIANTS.  
 RX MEDLINE=95038763; PubMed=7951253;  
 RA Roberts R.G., Gardner R.D., Bobrow M.;  
 RT "Searching for the 1 in 2,400,000: a review of dystrophin gene point  
 RT mutations";  
 RL Hum. Mutat. 4:1-11(1994).  
 RN (9)  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=94320940; PubMed=8045556;  
 RA Rimsland F., Reiss J.;  
 RT "Microlesions and polymorphisms in the Duchenne/Becker muscular  
 RT dystrophy gene.";  
 RL Hum. Genet. 94:111-116(1994).  
 RN (10)  
 RP VARIANTS DMD ARG-54.  
 RX MEDLINE=94004962; PubMed=8401582;  
 RA Prior T.W., Papp A.C., Snyder P.J., Burghes A.H.M., Bartolo C.,  
 RA Sedra M.S., Western L.M., Mendell J.R.;  
 RT "A nonsense mutation in the dystrophin gene in a Duchenne muscular  
 RT dystrophy patient.";  
 RL Nat. Genet. 4:357-360(1993).  
 RN (11)  
 RP VARIANTS DMD HIS-365; TRP-2191 AND ARG-2937.  
 RX MEDLINE=95352525; PubMed=7849724;  
 RA Nigro V., Nigro G., Esposito M.G., Comi L.I., Molinari A.M.,  
 RA Puca G.A., Pollano L.;  
 RT "Novel small mutations along the DMD/BMD gene associated with  
 RT different phenotypes";  
 RL Hum. Mol. Genet. 3:1967-1968(1994).  
 RN (12)  
 RP ALTERNATIVE SPLICING (DYSTROPHIN-1 AND -2).  
 RX TISSUE=Retina;  
 RA White R.A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN (13)  
 RP ALTERNATIVE SPLICING (DYSTROPHIN-3).  
 RX TISSUE=Brain;  
 RX MEDLINE=89181947; PubMed=2648158;  
 RA Feener C.A., Koenig M., Kunkel L.M.;  
 RT "Alternative splicing of human dystrophin mRNA generates isoforms at  
 RT the carboxy terminus";  
 RL Nature 338:509-511(1989).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE  
 CC PLASMA MEMBRANE.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING; THE 3 SHORTER VARIANTS ARE CALLED  
 CC DYSTROPHIN-1, -2 AND -3.  
 CC -1- DISEASE: Defects in DMD are the cause of Duchenne muscular  
 CC dystrophy (DMD) and Becker muscular dystrophy (BMD). DMD is the  
 CC most common form of muscular dystrophy; a sex-linked recessive  
 CC disorder. It typically presents in boys aged 3 to 7 year as  
 CC proximal muscle weakness causing waddling gait, toe-walking,  
 CC lordosis, frequent falls, and difficulty in standing up and  
 CC climbing up stairs. The pelvic girdle is affected first, then the  
 CC shoulder girdle. Progression is steady and most patients are  
 CC confined to a wheelchair by age of 10 or 12. Flexion contractures  
 CC and scoliosis ultimately occur. About 50% of patients have a lower  
 CC IQ than their genetic expectations would suggest. There is no  
 CC treatment. BMD resembles DMD in hereditary and clinical features  
 CC but is later in onset and more benign.  
 CC -1- DISEASE: Defects in DMD are a cause of X-linked dilated  
 CC cardiomyopathy (XLCM).

CC -1- MISCELLANEOUS: THE DMD GENE IS THE LARGEST KNOWN GENE IN HUMANS.  
 CC IT IS 2.4 MILLION BASE-PAIRS IN SIZE AND COMPRISES 79 EXONS.  
 CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY  
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIBRININ,  
 CC ABP-120, ABP-180, OR BETA-FODRIN).  
 CC -1- SIMILARITY: CONTAINS 2 CALPOTIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 MW DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.  
 CC -1- DATABASE: NAME=DMD; NOTE=Dystrophin Mutation Database;  
 CC WWW="http://www.dmd.nl/database.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 CC EMBL, X06179; CAA29545.1; ALT\_SEQ.  
 CC EMBL, X06178; CAA29544.1; -.  
 CC EMBL, X14298; CAA32479.1; ALT\_SEQ.  
 CC EMBL, X15495; CAA33518.1; -.  
 CC EMBL, X54820; CAA38589.1; -.  
 CC EMBL, X13045; CAA31451.1; -.  
 CC EMBL, X13046; CAA31452.1; -.  
 CC EMBL, X13047; CAA31453.1; -.  
 CC EMBL, X13048; CAA31454.1; -.  
 CC EMBL, U27203; AAA86115.1; -.  
 CC EMBL, X15148; CAA33245.1; -.  
 CC PIR, A27605; A27605.  
 CC PIR, A27162; A27162.  
 CC PIR, S05291; S05291.  
 CC HSSP, P46939; IQAG.  
 CC Genew; HGNC:2928; DMD.  
 CC MIM; 300377; -.  
 CC MIM; 310200; -.  
 CC MIM; 300376; -.  
 CC MIM; 302045; -.  
 CC InterPro; IPR001589; Actbind\_actuin.  
 CC InterPro; IPR001715; Calponin-like.  
 CC InterPro; IPR002017; Spectrin.  
 CC InterPro; IPR002349; MW.  
 CC InterPro; IPR001202; WW\_Rpds\_WMP.  
 CC InterPro; IPR000433; Znf\_ZZ.  
 CC Pfam; PF00307; CH; 2.  
 CC Pfam; PF00397; WW; 1.  
 CC Pfam; PF00435; spectrin; 22.  
 CC Pfam; PF00569; Z2; 1.  
 CC PRINTS; PR00403; WMDOMAIN.  
 CC SMART; SM00033; CH; 2.  
 CC SMART; SM00150; SPEC; 21.  
 CC SMART; SM00456; WW; 1.  
 CC SMART; SM00291; Znf\_ZZ; 1.  
 CC PROSITE; PS00019; ACTININ\_1; 1.  
 CC PROSITE; PS00020; ACTININ\_2; 1.  
 CC PROSITE; PS50021; CH; 2.  
 CC PROSITE; PS01159; WW\_DOMAIN\_1; 1.  
 CC PROSITE; PS50020; WW\_DOMAIN\_2; 1.  
 CC PROSITE; PS01357; ZF\_ZZ\_1; 1.  
 CC PROSITE; PS50135; ZF\_ZZ\_2; 1.  
 CC Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;  
 CC Repet; Disease mutation; Alternative splicing; Zinc-finger.  
 CC FT DOMAIN 1 240  
 CC FT DOMAIN 1 119  
 CC FT DOMAIN 134 237 CH 2.  
 CC FT REPEAT 339 447 SPECTRIN 1.  
 CC FT REPEAT 448 556 SPECTRIN 2.  
 CC FT REPEAT 559 667 SPECTRIN 3.  
 CC FT REPEAT 719 828 SPECTRIN 4.  
 CC FT REPEAT 830 934 SPECTRIN 5.

Query Match 2.0%; Score 213.5; DB 1; Length 3685;  
 Best Local Similarity 17.1%; Pred. No. 0.0013;  
 Matches 404; Conservative 347; Mismatches 830; Indels 777; Gaps 102;

12 EFRKTRALSKGTAAELAFQSVSEVVRGVLAKPELLEPLDYENVYQKTOIINDCIR 71  
 326 EKSFGSSLSMESEVULDRYOTALREVLWLSAEITL-----QAQGEINDV-- 372  
 72 EWL-----LPEYDFQFOAILLRGRY-----ICSTVPAKAEEBAQSLFYRECIKTNS 119  
 373 EVVKQGFHTHEGYMDLTAHQGVGNILQSGKLIGTGKLSEDETE---VOEQWMLNS 429  
 120 DWHLVNY-----KYEDYSGEFRQLPNKVVLDTLPHYVVEVDEVDKED---AASIGSQK 172  
 430 RRECLRVASMEKQSLHRYLMDLQNG--KLKELNMDLTKTEERTKMESEPLGPRLEDK 487  
 173 GGTTHGMYKGNMNSAISVTMRSEFRPFHLIQLDGSYKPEFLDKLQEKPGSIPLGF 232  
 488 RQVOCHKVLQEDLEQEOVRVNSLT-----HMVVVDESSGDHATVALLSEQLK----- 534  
 223 LYGVSFRNNKVRRAFAELKMDKSSVLLADSEVEHEWITLNLKIL-----QLNPE 284  
 535 VLGDWMAN--ICRMT-----EDRWVLLODILLKMQRLTEBQCLFS 572  
 285 AAMQEKRGDSDHEDEQSKLEGG--SGLDVYLPDLAKSA-REAEIKLKSEHVKLFYLDP 342  
 573 AMLSEK-----EDAVNKIHTTGFQDQNMELSSLOKLAIVLKADLEKKQSGKLYSLKQ 625  
 343 D-AQDLDPGSAPEVKSFEKFGKILVKC-NDLSFNIQCCVAENEDEGTTNVEFPVTL 400  
 626 DLSTLTKMKSIVQKTEAWLDF-----ACMDNLVQKLEKSTQAQISQAVTT--TOPSLNQT 679  
 401 SLFD-----IKY-----NRKLSADFHV---DLNHSFRQMIATTS 432  
 660 TMEVTVTTTREQILLVKHAELEPPPPQKRRQTLTVESEIKRRLDVTDELHSMI---- 735  
 433 PALMNGSGPETQSALRGILHEAMQYPKQGISVTCPHPI--FLVAREIKVQSGSITHCA 491  
 726 -----TRS-----EAVLOSPEFAIRFKSGNFSULKEKNALIER-----EKA 771  
 492 EPMYKSSDSSKVAQKVLK-----NAKQACORLQGYRMPFMAAATFLFKASGND 541  
 772 EKFRKLQASRSQAQLVEQMGNEGVNADSIKQASEQLNKHWFEC---QLSERLNMLE 827  
 542 KNARSATYRQ-----DSNKLSDNDMLKLADFRKEBKAKLVEYLGNDDITIDNVS--SD 595  
 828 YQNNIIAFNQLOOLEQMTTAEMLKI-----QPTTPEPTPAIKSOLKICDEVNRLSG 882  
 596 PPNYVNSSYIPTKQFETCSKTPITEFEVEEFVPCIPKHTQPYTITNHLVVPKYLKYSQ 655  
 883 LQQLERLKTQSIALKEKGQPMFLDAD-----FYAFNNHF---KQVPSDVQ 926  
 656 KSFAPARNIACIEFKDSDDEDSOPLKCIYGRPGGVPFTSAPAAVLHNNQPEFYDEIK 715  
 927 AREKELQITF-----DTLP-----PMRYQSTMSAIRTWVQO-----SETK 961  
 716 IELPQLHEKHHLLTFPHVSCDSSKSGTKGRDVEYTOVGYSMPLKDGAVVTSEQHI 775  
 962 LSIPL-----QLASTYDYEIMEORGELOALQSSLOEQSGLYYL-----STYVK 1004  
 776 PVSANLPGCYLGELOELMGRAHYGPEIKWVDGKPLIKTISTHLVSTVYTTQDHLNHPQVC 835  
 1005 EMSKAPRS-----EISRKTKQSEPEIEG--RMKULSSQV-----EHC 1040  
 836 QKTESGAQALGNELVYKLSLHAMEGHVIAFLPTLLNQLFRVLTTRAQOEVAVAVNTRYI 895  
 1041 QLEF-----EQNNKLKRIQWHIQ-----TLKWMMAEVDVFLKEWPAIDGSEITL 1084  
 896 IHVAVQCH-----EGLGESHLSRYKAVKPEPYASEYKT----- 931  
 1085 KQQLQCHLLVSDIQTIQPSLNSVNEGQKIKINEABEPFASRLTELKELINTQMDHMCQ 1144

932 -----VHEELTKSMTTLIKPSADPLTSLNULKYSMFFFDVL 967  
 1145 VYARKEALKGGLKLEKTVSLQKOLSEHHEMTQAESEYLERDREYKTPDLQK----AVDEM 1200  
 968 IKSMAGHLENSKYKLYLRNQPASYNHNAVETVNMMLPHLTQKFRDPEA-SKAN--- 1023  
 1201 KRAREBAQKAKAYLTL-----ESVNSVIAQAPPAQOELKSLLETLTNNYOWLC 1251  
 1024 -----HSLAVFIKRCPTFMDR--GFVFKQJNNYISCAPODKTLFPEYK 1065  
 1252 TRLNCKCTLEBWACHHELISYLEKANKMLNEVEFKLTEN-----IPGAEISSEV- 1305  
 1066 FEFLRVNCHRYIPLNMPFPGKRIQRYODOLDVSLTDEFCRNHFLVGLLEEVGTA 1125  
 1306 LDSLLENLRHSEDDN-----NQRIILAQTITDGVMD-----LINELETFP 1347  
 1126 LOEFREVLIAISVKNLLIHSFDPRVASHSHQARIYLYPLFGLLEINVQRIWADV 1185  
 1348 NSRMELEHEBAVRQK-----LLEQSIQSAOETEK 1377  
 1186 SPPFNAGMTVKDSIA---LPVNPVLTPOKSGTLDNSLHKDLGALSGIASPYTSTP 1242  
 1378 SLHLIQESLTIIDKQLAAYIADKVDAAQMPQOAKIQSDL----- 1417  
 1243 NINSVRNADSRGSLISTDGSNLSLPERNSEKNSLIDKHOQSTLGNVYRCDKLDQSEIKS 1302  
 1418 -----TSEHISLEE-----KKHNGQGBAQRVLSDQDVQKQLQD 1453  
 1303 LMKCFLYLKSMSDALTFTYNNKASTSELMDFITISEVCLHOFQYMKRYIARTGMHAR 1362  
 1454 VSM-----KFRLLFOKRPANFEL-----RLQESKMLIDBVXOHLP 1486  
 1363 LQQLGLDLSLT---FNNSYGHSDADVLHOSLLEANITAEVCTALDLSLFTLAFNQOL 1419  
 1487 ALKTSYQOEYVQSOQLNCVN-----LYKLSLSE--VASEV-----EM 1521  
 1420 LADHGNELMKKVPDVVLCFLQKQSE-----TALKRVFTALPSLIYKPFSTPYEG 1470  
 1522 VIKGRQ-----IYQKKQTEPKELDERVNTALKHMYELGAKVTERQ----- 1564  
 1471 RADMCALCYEILCKCSKLS--SIRTEASQLLYFLMRNNPYTGKGFV----- 1518  
 1565 -----OLEKCL--KLSRKQRKEKNVLTWELAAITDMELT--KRSAGEWPSNLDSBV 1611  
 1519 -----RTHLOVITISVQGLADVVIGIGTRQGSLSIINH-----CANSDBLI 1560  
 1612 AMGKATQKEIEKQKVNHLSITEVEBALCTVAGKKEITVEDLKLSLNSMIAVTSABEWL 1671  
 1561 -----KH--TSFSSDVLDLTKRI--RTVLATAQOMKEHNDPEMLVDLOYSLAKSVAST 1610  
 1672 NULLEYQGHMETFQONVDHITKWIIOADPTLDBSEKKGPOQKEVDVLRKALNDIRKIV 1731  
 1611 PELRKTWLDNARIHVKNGLDSEAMCYVHTALVAEYLTRKEAVOWEPPL--LPHSHA 1668  
 1732 DSTR-----DQANIMANRGD-----HCRKLV-----BPOISELNHRPAA 1766  
 1669 CLRSRGVFRQGCATAPRVTIPNDE-----EASMEDVGMQDVHFNPDV-----LME 1716  
 1767 ISHRIKTGKASIPKLEBQFNDSIQKLEPLEASIQOVNLEKEDFNKDNMNEGTYKE 1826  
 1717 LLEQCADGLMAERYELIADLYKLIIPYERKRPFERLAHLYDTLHRAV-----SKV 1768  
 1827 LLQR--GDVL-----QQRITD-----ERGREIKIKOULLQTKHNAKULRSORRKA 1872  
 1769 TEVMSGRRLLTGYFRAVFGQAQOYQFTSETVDEGEFEDEDEGEYIYKSPK-LTPUSE 1827  
 1873 LEISHQ-----WYQYKQADDLKLCLDIDIEKKTASLPEPRDERKIKE 1914  
 1828 ISQRLK-----LYSDKRGSEVNMKIQDSGKNP-----KOLDSYAYIQTHTVIRP 1874  
 1915 IDRELQKKKEELNVRQAEGLSBDGAAMAVEPQIOUSKRWIRIESKFA----- 1964  
 1875 FDEKELQERKTEPFRSHNIRRFMEFMPPTQYKQGGVEQCKKRTILTALHCPYVKKR 1934



```
Db 1152 --KILLSODAKTD-----KIVLVN-----THLLKIKNTQFETAPAKALTVBEV 1195
Qy 1797 TDSFVDSGFEDDEGKEIYKPLTLPSEISORLLKYNSKPSSENMIDSGKNP 1856
Db 1196 NKNKQQL-----QELLNHF-----NAHLKAMDDNNHTK 1224
Qy 1857 KDLDSKY-AVIGOVTV-IPFDEKEIQERKTEFER-SHNIIRFMEFMTGTGRQGGV 1912
Db 1225 LSPDTKMTAFVSDSLYIPDFINIKLQDPSDPFVISOGLMKANEMPY-----I 1275
Qy 1913 EEOCKRRITLVAIHCPPIYKRI PMVYOHNDLNPLEV-AIDEMSKV-----AELRQIC 1966
Db 1276 QAEITLKMLTKLIH-----DINKFCLSAISEFGKEALPFVYALRLDE 1318
Qy 1967 SSAEDWIKLKQKJGSGSVSVQVNAGBLAVARAFLDQNTKRYDPDKVKLLKEVFQFVEA 2026
Db 1319 VOINTKVEIEKNVICNETVE-----NTKNIP-KLTKLLKQLVPRKV-A 1360
Qy 2027 CGQ--AAVNERLIKEDQLEY---QEEKANRYREMAKELSEI-----MHEQICPLE-- 2072
Db 1361 GGOBOYQVTLNKNILSESTSMQOTYERKQKEYFEIYVNVNVAFKLAFFNPQQLQVVERLI 1420
Qy 2073 EKTSLVLPNS 2081
Db 1421 EKFKSLPKRS 1429

RESULT 3
CENP_HUMAN STANDARD; PRT; 3210 AA.
AC P49454; Q13246; Q13171;
ID 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yan T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN (3)
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=9536446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN (4)
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN (5)
```

```
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-B reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION. INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
CC -1- SUBUNIT: HOMO- OR HETERODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),
CC REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF
CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
CC -1- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
CC -1- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; U19769; AAA82889.1; -
DR EMBL; U30872; AAA82935.1; -
DR EMBL; U25725; AAA86889.1; -
DR Genew; HGNC:1857; CENPF.
DR MIM; 600236; -
DR InterPro; IPR001230; Prenyl_site.
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1328
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT REPEAT 3015 3032
FT DOMAIN 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561
SQ SEQUENCE 3210 AA; 367589 MW; 11D8332496084334 CRC64;
Query Match 1.8%; Score 199; DB 1; Length 3210;
Best local similarity 18.5%; Pred. No. 0.0074;
Matches 424; Conservative 342; Mismatches 828; Indels 700; Gaps 107;
Qy 94 CSTVPKAEBAQS-----LFVTECIKTYNSDMLVNYKYEDYSGEFDLPN 140
Db 121 CKSELERQQAQASDVSLNCPPTPKFTPLPFSQ-----YSGSKYEDLKERY-----N 172
Qy 141 KVVKLKDKLPVAVYDEVDDDAASISGSKG-----ITKH-----GWL 181
Db 173 K-----EVEERKRLAEAVKALQAKKASQTLPOATMNRRLIARHQASSSVFSQW 220
Qy 182 YKGNMNSAIVTMSFKRRFPHLIQLGDSYKFFELKQLQKRPK-----GSIFL 230
Db 221 QEKTPSHSSNSQTPPIRDF-----SASV-----FSGELVTPRRTQIGKRDANSFF 271
Qy 221 G-----FLYGVSRNNKVRRAFELKMQDKSSVYLAADSEVEMEWITLANKI--IQLN 282
```

Db 272 GNSSSPHLLDQKAOQONELRNKINLELR-----LOGHEKEMKQ-----VANKFOELOQ 321  
Qy 283 PEAAOKEKNGSDHEHDEQSGKLEGGSGLDSPYLPAKASR--BAIKTKS-ESRYKLE 338  
Db 322 LEAKKEL-----IEKEV-----LANKDELVRTTAQYDQASTYTALEQTKL 367  
Qy 339 YLDPAQKLDPSA-----EPYKSFEEKFKRILVKNLSPNLQCCVANEBCPTTAYE 394  
Db 368 TEDLSGONASGASCSLEOKIKEKEKEFOEL-----SRQO 404  
Qy 395 PPRVTLSTFDIKNRKISADFNVDLNFHFSVRO-----MIATSPALMNGSGPETQSLRGI 450  
Db 405 RSFOITDQOCIQKAKLTQELQAKMHNVLQAEILDKITSVKQOLNN----- 452  
Qy 451 LHEAAMQPKQIFSVTCPPHPIFLVARI-EKVLQSGITHC-ABPYKSSDSKQAKY- 507  
Db 453 -----LEEFKQK-----CBAEQAFQASQIKENELRSMEMKENLLKSHSGQKAREVC 503  
Qy 508 -----LKNAKQA-----CQRLGQYRMPRMAAFILFKDASGNDKNARFSAIYRQDSNKLSN 559  
Db 504 HLEAELKNIKOCUNQONFAEMKAKNTSOETLRLQEKIN-----QOENSLTL 553  
Qy 560 DMLKLLADPRK-----PERKATPVILGNLDTITNVSSDPYNYVNSSYIPTKQETOS 614  
Db 554 EKTLAVALDEKQKQDCSODLLKREHNIQLNDKLSKTERESKALLSLELKKEKEBELK 613  
Qy 615 KTRPTE-----EVEEFVPCIPKHTQPYTITNHLVYPR-----YLYKQSG 655  
Db 614 EERTLPSCKSENEKLLTQMESEKENLOSKINLEFLCTQOIKSHENYERVTLEMDIE 673  
Qy 656 KSPAKARNIACIEFKDSDEBDSQPLKCIYGRGQVFTRSAFAVLLHHQNEFYD--- 712  
Db 674 NLSVEIRNLNHLVDSK-SVEVEFOKL-----AYMELQKAEFSDQGH 714  
Qy 713 --EIK-----IELPQHEKHLLTLTFHVSCDNSKSGSTKRDVVEQVGSYMLPL- 763  
Db 715 QKEIEMCKTQSQTQVEDLEHKLQL-----SNEIMDKORCYODLHAEBESLRDL 767  
Qy 764 --KQGVUTSEOHIPVSAULPSGYLGQELGMRHYPREIKWQDGKPLIKISTHLVSTV 821  
Db 768 KSKDASLVINEDI-----QSLIAFDQOPAMHNSFANIIQEGSMPERSECRLEA- 818  
Qy 822 YQODQH-----LHNFQYCOQTESGAQALGELVYKLSLHAMEGHVIAPLPTLNLQ 875  
Db 819 ---DQPKNSAIIQNNVDSLEFLESQKQMNSDIQKQCELVQIKEIEENLMAK--EQM 873  
Qy 876 FVLVTRATQOEVAVNVTRVLIH--VVAQCHEBGLJEHLRSYKAKAPYASEVKTYN 933  
Db 874 HOSFVAETSORISKLOEDTSAHQNVAA-----TJSALENKEKELQLNDKRVETEQAET- 927  
Qy 934 BELTKSMITILKPSADPLTSNKLKYSWFFDVILKISMAQHLENSKVLLRNQRPASY 993  
Db 928 QELKK-----SNHLEDSLKELOLSTLSTLEKKEKMSIISLNR- 968  
Qy 994 HHALETVMMLPHITQKFRDNPEASKNANSLAVFIKRCFTFMORGFVKOINNYISCF 1053  
Db 969 -----IEELTOENGTLKEINASINOEKMLIQKSESFA-----NYIDER 1007  
Qy 1054 AFGDPKTLPEYKFEELRVVCNHNHYIPLMLPMPFGKRIOR-----YODLODYSLTD 1106  
Db 1008 EKXISISLQYKQEKIL-----LORCEETGNAVEDLSQKKAQO 1047  
Qy 1107 BECRNHFVLGLLEVGTALEFREVRLLAI SVLKUL--IKHSPDRYASRSHQRIAT 1164  
Db 1048 E---KSKKECLINEC-TJSLCENR-----KNLEOLKEAF-----AKEHGFELTK 1088  
Qy 1165 LYLPLFG-----LLIENQRIINVDSPPFVN---AG-----MTVKQESIALPAV 1206  
Db 1089 L---AFAEBRNQVLMLETVQOALRSEMTDNONNSKSEAGGKQEIIMTLKE- 1138  
Qy 1207 NBLVTPQKSGTLDNSLHK--DLGAI SGIASPYTT-----SPNINSTRNADSRGS 1255  
Db 1139 -----ONKQKEVNDLLOENQOLMYMKTKHECONLSEPIPINSVYKERSERN 1186

Qy 1256 L-----ISTDSGNS-----LPERNEKSNSLDKHQOSSTLGN-S 1288  
Db 1187 QCNFKRQMDLEVKELSLDSYNAGOLVQLEAMLRNKLKQSEBEKECEQHLEQTRGLE 1246  
Qy 1289 VVRCMDLQOSEIKSLMKCPL-----YI-----LKMSDDLFTYMNKASTSELMDFPTI 1337  
Db 1247 TSNLODMQSOEISGLKDEIDAEEKYIGSPHELSQDNMHLQCSLOTWMMKINLEKI 1306  
Qy 1338 SEVCJHQYQWYMKRIARTGMNAR-----LOOLGSLNLSLTFNHSQSDADVL 1387  
Db 1307 CBI-----LOAEKELYELNDSRSECTIATRKMAEBVGKLINEVKIL---NDSGUL 1356  
Qy 1388 HQSLLE-----ANATEVCCTALD-----TJS-----LFTLAF 1415  
Db 1357 HGEIYEDLPGBEFGRQPHQHPVSLAPLDEGNSYEHLLTSDQEVQMHFAEIOEKTLQSG 1416  
Qy 1416 KNQIADHGHNPMLKQVFDVYLCEIQAQSEFALKNVF--TALRS---LYKFPSTYE 1469  
Db 1417 EHKILHDO-HCOMSKMBE-----LOTVVDSLAKENLVSLTLNRFQGDVLVEMQGLBE 1470  
Qy 1470 GRADWCALCYEILKCSKLSIRTEASQULLYFLMRNPNPYTGKSPVTRHLOVITISVS 1529  
Db 1471 GLVPSLSSCVB---DSSSLSSLGDS-----FYRLLEQTDGMSIL-SNLGAVSAN 1519  
Qy 1530 QLIADVIGIGTRFQOGLSIINNCANSDRILIKHTSFSSDVLDLTRIRTVLMATAQMKH 1589  
Db 1520 QCSVDEV-----FSSLOTYVDSLAKENLVSLTLNRFQGDVLV- 1558  
Qy 1590 ENDPMLVDLOYSLAKSYASTPELRYTLDSMA-----RHHVKNQGLS----- 1632  
Db 1559 ---EMQGLBEGLVPSLSSSCVDPDSSLSLSSGSPFYRALBQTDGMSLSSNLGAVSA 1614  
Qy 1653 --BAAMCVHATVLAVEYLTRKBAVQWMBPRLPHSHSACLRRSGRGVFQOGCAFPVITPN 1691  
Db 1615 NOCSYDEVFCSLQEBNLTREKTPESAPRAKVYBELSLC-----EYVRQ-----S 1658  
Qy 1692 IDEBASMEDEVQMDVHFNEDVLMELLEBOCADGLMKARVEL--IADYKULIIPYERKR 1749  
Db 1659 LEKLEEKESQGIK--NKEI--QELEQ-----LSSRQELDCRKQY---LSEBQ 1704  
Qy 1750 DFERLAHL---YDTLHRAYSKYTEVMH-----SGRLLGTYFRVAVFGQA 1791  
Db 1705 WQCKLTSYTLMEESKLALEKKQTEQLSLEVARLOLOGLDLSRSRLG- 1753  
Qy 1792 AQOFTSETVQEGFDED--GKEYIYKPLTLPSEISORLLKYSDKFSB- 1846  
Db 1754 ---IDTEDAIQGNESCDISKEHTSETTERPHGVH-ICCKDAQOOLNLDIE 1803  
Qy 1847 MIDSGKYNP-----KOLDSKYAYIQVTHVIPFEDEKELOERKTEPERSHNIRFME 1899  
Db 1804 KITERGANVKTPEGSGEGSDPTNVE-----PRGEDKTQGSSECSISELSPGNALVP 1855  
Qy 1900 MFTQGTGRQGVQECQKRTITLAIHCFPYVKRIPIPMYQHTDLDNIEVALDEMSKV 1959  
Db 1856 MDFL-----GNOED-----IH-----NIQLRYKETSNEMLRLIHHV-IEDRDRKV 1893  
Qy 1960 A-----ELRQCSSADEVMIKLOLKQGSVSQVANNAPLAVARAFDINTKRYPNKYKL 2015  
Db 1894 ESLNMEKELDSKHLQEOVLMTKIBACTIELEKVLGELKKNSDSE-KLEFSCDHQEL 1952  
Qy 2016 LKEVRFQVEAQQALAVN-ERLIKED-----OLEYQEBKANYREMAKELSEIMHEQICP 2070  
Db 1953 LQRV--ETSSEGNLSLEHNAHAKSRREDIGDVAVANDSKERFLDVENEISRIR----- 2004  
Qy 2071 LBEKTSVLPNSLHT 2084  
Db 2005 -SEKASIBHAYL 2017

RESULT 4  
NM1 YEAST ID NM1 YEAST STANDARD; PRT; 2748 AA.







```

Qy 897 HV--VAQCHBEGLESHLSRYKAYAEYVASEYKTHAE--ELTKSMITLIKPSADF-- 950
Db 261 HAPHRKKNKSHSLLELENDKKKDDTASTLYVATEVDLIPSEYELNSQFSLPHDKSPFI 320
Qy 951 LSNKLKSKMFEFVDLISMAQHLENSKVLBNQRPAPSYHAAVEVVMMLMHPHQ 1010
Db 321 MKN--FNWITY--LQDSPIN--PQOPATPP--LNNRVHLLCOIID 360
Qy 1011 KF-----RDNPEASKNANSHLAVIKRCFTFMDRGFEVFKO-----INNYS 1051
Db 361 VVGVSSTNDSSKQOQOT--IFINYLRFKDPSEFVFNALAACQENNIDLIYINNYL- 417
Qy 1052 CFAPDPTLFEBYKE-----FLRVNC--HENY-----IPLNMPFSGRIQRY 1096
Db 418 C-----KTTFRTERILLSKFLAVDNDHKHKEVWKSATTQMFQPKLDIENYL 471
Qy 1097 DLQDLYSLDER-----CRN--HFLVGLLREVGTLQDFEVRRLAISVLLKLIKSP 1149
Db 472 KAYVHNPNVHHEFLCLNKAERKYVAVLNKRKEIOAIERKNKSFQLSN-----F 525
Qy 1150 DRYASRSHQARIATLYLPLFGLLENVORINRVDSPPPVNAGMTVKDESLALPVPNDL 1209
Db 526 IDKGE-----TPALPIESENVAKHTSDTBEIGIVERPFI--ESTIELPISTL 571
Qy 1210 VTPQKSTLNSLHKLGLGASPYTST--PNINSVRNADSRGSLISTD----- 1260
Db 572 NNTQOIS--LDKQINELTWTITHTLTKFTKIVODNNNI--AAGFMVTELCLEFAYL 626
Qy 1261 -----SGNSLPERNSEKNSLDKHOOSTLNGSVVRCOKLDSKLSLMCL 1308
Db 627 VNLVYNIEVLKSHGSLNITVLLQEVKLDNDQFLRFGSHNNINLSN-----FT 677
Qy 1309 YLIKMSDPLFTYNNKASTSELMDFETISEVCLHQFQVWG--KRYIARTGMMHARLO 1364
Db 678 LSRKMFVU--FINSOKPSD--RASEILAEISLADPSKKNYVNIEMIKSOIE 728
Qy 1365 QUGSLDNSIJTFNHS-----YGHSDADVLHQSLELANIATEVCLTALDLSLFLAK 1416
Db 729 ELGKKEISTTENKQFALTKQILGDDELPIYDFLHLSAVALPNTTYKNLHLHLEIQ 788
Qy 1417 NOLLADHGINPLMKKVPDYLCFLOHGEFALKNVFTLRSLIYFEPSTFPEGRAOMKA 1476
Db 789 RPDIMATLHDKI--QSLDIYVDMT--NDITYEQTFSTVLVDLFPNSTEK-----T 839
Qy 1477 ALCYEIL-----KCCNSK-----LSISRTESQULYFLMRNNDFDTGKKSFRTHL 1522
Db 840 ALPESVLTIRQAKKCANIKTIDEPDLOFTTMSQONNRQHF--GKK----- 889
Qy 1523 QVIVSOSLIADVIGIGTRFOOSLSIINNCAUSDRLIHTSFSSDVKDLTRIRITVL-- 1580
Db 890 -----IPLT-----MGHIKFLYSOKI-----TTEKKWIGRAKTVVIT 923
Qy 1581 -----MATL-----OMKE--HENDPEMLVDLOYSLAK-- 1605
Db 924 SPEELTAFATAPTKALQOTCKPDLDKALQRMHEEMOKJAENDKXIIITISLTLEKRL 983
Qy 1606 -----SYASTP-----ELAKTWL-----DSMARITHVKNGLSEAAVCYVATLVMEY 1648
Db 984 DILLILKQGFSSLETMLNLLETFLKQLODNNVNIHFTALLPVLKDIETTSKISIV 1043
Qy 1649 LTRKAVOWEPPLLPKSHS-----ACLRSRGCVF-----ROGCTAFR 1686
Db 1044 I--EKILIKTLPNPKONSKEOKYTPILSPSKFKTTCFEDVTEIEYOQKSTPLK 1100
Qy 1687 VITPNIDEASMEVDGMDVHFNEDVLMEL-----LEQCADGIMKEREYELIADIYK 1739
Db 1101 KATSTNKHTRLSHSYGQELWLYERITELKKTNNKKEOJSKEVALAEK-----K 1152
Qy 1740 LIPIYERKDERELANLVDTLHRAVSKTEVWHSGRLLGTGYFRVAFQAOYQFPTS 1799
Db 1153 ILLSODAKTD--KIYLVAN-----THIKKIKTQFKETAPAKLTYEVANNK 1198
Qy 1800 ETDVEGFEDDEKKEYIKPLTPLSEISORLLKLYSDKFSSENVKMIQDSGKVPKDL 1859

```

```

Db 1199 ENQI-----OELNHF-----NALKKNMOPNHTKLSF 1227
Qy 1860 DSKY-AVIQVHV--IPFDEKLOERKTEFER--SHNIRRFMFEPFOTGRQGVBEQ 1915
Db 1228 DTKMTAFVSDSLYIPDINIKLOPISDPFKVLSQLMKNTNEMP-----IQNE 1278
Qy 1916 CKRRITLTAHCFPVYKRIIPVMOHTDLNPIEV-AIDENSKV-----AELROLCSGA 1969
Db 1279 ITLKMLTLIH-----DINKFCLSAISFGKEMIPNVAALRLEVOI 1321
Qy 1970 EVDMIKLOLKQGSVSVONNAGPLAVARAFDDNTTKYPPNNKYLKVEPRQVYACQ 2029
Db 1322 NTKVYIENKVIENCTEVE-----NTKNIP-KLTKLKLQDLPKRY-AGSQ 1363
Qy 2030 ---ALAVNERLIKEDQLEY--QEBKANYREMAKELSEI-----MHSOICPLE--EKT 2075
Db 1364 EOYQTLMMKILITSETSMOQTEKEQLKREYBEIVNVASFPLAENFPQOLQNVRLIEKP 1423
Qy 2076 SVLPNS 2081
Db 1424 KSLPKS 1429

RESULT 6
G160 MOUSE STANDARD; PRT; 1325 AA.
ID G160 MOUSE
AC P55937;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Golgin-160 (Male-enhanced antigen-2) (MBA-2).
OS GOLG3 OR MEB2.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Testis;
RX MEDLINE=97217683; PubMed=9063644;
RA Kondo M., Suto S.;
RT "Cloning and molecular characterization of cDNA encoding a mouse
RT male-enhanced antigen-2 (Mba-2): a putative family of the Golgi
RT autocatalign";
RL DNA Seq. 7:71-82(1997).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
CC DETECTABLE MALE ANTIGEN (SDM).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
CC LEYDIG CELLS, SPERMATOCONIA, OR SPERMATOCYTES.
CC -1- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-19 OR MET-30 IS THE
CC INITIATOR.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; D78270; BA019612.1; -.
DR MGD; MGI:96958; Golga3.
KW Spermatogenesis; Developmental protein.
FT DOMAIN 201 204 POLY-ALA.
SQ SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;

Query Match 1.8%; Score 193; DB 1; Length 1325;
Best Local Similarity 19.2%; Pred. No. 0.0042;
Matches 191; Conservative 172; Mismatches 389; Indels 242; Gaps 46;

```

```

QY 1215 GSTLD---NSIMKDLGAIISGASPYTSTPNINVRNADSRGSLISTDSGNSLPERNGE 1271
DB 99 GSTADAVGSLKNS--RSSTSVSEVSPSESTDNRESASMTGDSVSEADGN---ESDS 153
QY 1272 KSNLDKHOOSSTLGN-----SVRCDKLDQ-SEIKSLIMCELY----- 1309
DB 154 SHSLSLARGACVGLVGNMGPTAYMVGQEISALGQPPSIDVLAQAAAHQHDQNOBA 213
QY 1310 --ILKMSDPAALTYNNKASTSELMPF-----TISEVCLHQFQWKGKVIAR 1355
DB 214 NGEVRNRDRSICSSVSMESSELAPODELLQIKKRRLEGVAFSLASQALQEK----- 269
QY 1356 TGMWHAARLQ-QLGSLDNLSTFNHSGH-----SPADVJHOSLBNATEVCLTA 1404
DB 270 ----AELQAQLAALSTRLOAQVHSHSQKODSLSEVDTLKQSCMDLGRMTDLOQM 324
QY 1405 LDTLSLFTLAFKQOL-LADHGNPLMKVFDVYLCLQKHQSESTALKVNFALSLIYK- 1462
DB 325 LEKNAASLASNNDLQVAEEQYRLMAKVEDMORNLTKONTVHDLRQOMTALQSOLOQV 384
QY 1463 -----FPSTFYGRADMCAALCYELLKCNKSLSTRBASQLYFLMRNPFDTYKGS 1516
DB 385 QLERTTLTSKLQASQAEI-----TSLQHARQWYQOQLTLAQBARYRLQGENA 431
QY 1517 FVRT-----HIQV-IISVSOLIADVIGIGTRFOQSLIINNANSDDL-IKHS 1564
DB 432 HIQVQMTQAGLEHLKLENVSLSHQTLER-----QHRI-----KEKELIAQLOS 478
QY 1565 PSSDVQDLTKRIRTVLMATQMKENHNDPMLVDLOYSLAKSYASTPELRKTWLDMA- 1622
DB 479 IEADMDOEAFAVQIREAKTWEE-----DLORLREEFEGEREQLOKV-ADAAAL 528
QY 1623 -----RIHVKNKGDLSEAMCVVHTALVAEVLTRKEAVQMPRLPHSHSCLARRSG 1675
DB 529 EQGLEOVKTLFRDQDLALQOEHLQD-VIKULTSTQEBALQAGSLDDLH----- 578
QY 1676 GVFRQCTAFRVTPINIDE---EASNMEDVGMQDVHF--NEDVJLMBEQCAGLWKAER 1730
DB 579 -----TRYDELQARLELQREADSRDA---HFLQNEKIVLEVALQSA---KSKK 623
QY 1731 YELIADYKLIPIYKRRDPFERLAHLVDTLHRAYS-KTEVMSGRRLIGTYFRVAFRG 1789
DB 624 BEIDRGARRL-----EEDTETSGLLEQLRDLAVKSNQVHLOQETATLRKQM----- 672
QY 1790 QAQVQFTSETDVEGFEDDEGKEYIYKEPKLT-----PLSEISORLLKLYSDK--FG 1841
DB 673 QKXKEQFVOQKQWVEAYRRDAYSKDOLINELKATKRLDSEMKELQOBLIKLQGEKTYE 732
QY 1842 SENVKMIODSGKNPK--DLSKVAIYQVTHVLPFDEKELQERKTEF-----ERS 1890
DB 733 VHSRRLQKMSLVHQMALEHGLSQVQKER-----DEMEIHQSLKPFDEKQMIALTEN 787
QY 1891 HNIRRFEMFPTQTSKRGQGVBEQCKRRTI--LTAHCFPVYKRIPIVMTQHTDLP 1947
DB 788 ETLKXQIEEL---QOEAKKAITQOKMKRLGSDLTS-----AQEMTKTKHAY----- 833
QY 1946 IEVAIDEMKVAELQLOCSAEVMDIKQLKQSVSVQVNAGPLAYARAPLDDNTK 2007
DB 834 -ENAAVSILSRLOE--ALASKENTDLELNQLRQO---STGSSDPVLAKIRALELELON 887
QY 2008 YPDNKKLLEKVEVQFVEACGALANVERLIKEDOLEYOBEMKAN--YREMAKELS----- 2061
DB 888 VQGSKLLEKEL-QEYITMTSOLERS-----REKYLEDELQESGFRKRIKRLBSNK 942
QY 2062 ----BIMHEQ-----ICPLEKTSVLPSNL 2082
DB 943 KLALEHERGKLTGLGOSNAALREHNSILETAL 976

```

RESULT 7  
DHC\_CABEL  
ID DYC\_CABEL STANDARD; PRT; 4568 AA.

```

AC Q19020;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dynein heavy chain, cytosolic (DYHC).
GN DHC-1 OR T21E12.4.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RX MEDLINE=96114101; PubMed=8674131;
RA Lye R.J., Wilson R.K., Waterston R.H.;
RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
RT nematode Caenorhabditis elegans."
RL Cell Motil. Cytoskeleton 32:26-36(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Maggi L.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC - SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, L33260; AAC37251.1; -.
DR EMBL, U80440; AAK21472.1; -.
DR WormPeP, T21E12.4; CE23997.
DR InterPro, IPR003593; AAA_ATPase.
DR InterPro, IPR004273; Dynein_heavy.
DR Pfam, PF03028; Dynein_heavy; 1.
DR SMART, SM00382; AAA; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 587 652
FT COILED COIL (POTENTIAL).
FT DOMAIN 814 844
FT COILED COIL (POTENTIAL).
FT DOMAIN 1241 1274
FT COILED COIL (POTENTIAL).
FT DOMAIN 1324 1340
FT COILED COIL (POTENTIAL).
FT DOMAIN 1559 1591
FT COILED COIL (POTENTIAL).
FT DOMAIN 1964 1992
FT MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3132 3229
FT COILED COIL (POTENTIAL).
FT DOMAIN 3339 3432
FT COILED COIL (POTENTIAL).
FT DOMAIN 3707 3739
FT COILED COIL (POTENTIAL).
FT DOMAIN 4359 4386
FT COILED COIL (POTENTIAL).
FT NP_BIND 1865 1872
FT ATP (POTENTIAL).
FT NP_BIND 2163 2170
FT ATP (POTENTIAL).
FT NP_BIND 2537 2544
FT ATP (POTENTIAL).
FT NP_BIND 2880 2887
FT ATP (POTENTIAL).
SQ SEQUENCE 4568 AA; 521568 MW; 028B52684F381676 CRC64;
Query Match 1.7%; Score 190.5; DB 1; Length 4568;
Best Local Similarity 18.0%; Pred. No. 0.039;
Matches 409; Conservative 319; Mismatches 798; Indels 741; Gaps 114;
QY 151 HYEVEDEVDKDEDAAISGQKGG-----ITRGWLKGNMNSAIVTMRSFRRRPFH 203

```

Db 66 NVIVDVRVAREOGQOPAGESGSEBAPATFOVHDLFMTRGQAMFVKOS-----N 119  
Qy 204 L10-----LGDGS--YKFEPLKDLQKEKSGIPLGFLYGVFPNNKYRFAF 248  
Db 120 V1EAEKXIAQVSAFPLNGSAMQOHLFMSRLNLNYCKS-7IG-QSGREBERODKLAFAV 178  
Qy 249 ELKMODKSYLLAADSEVEMEWITLILKILQINFEAMQEKRN-----GSHDEDEOS 302  
Db 179 QOCFTEAEALHLQONIDIPRINVLVNOHILDAIQAGKENBRAKIEDIGLVED--AN 236  
Qy 303 KLEGGSSGIDSYLPELAKAR--EAEIKLSSESIVKFLYDPPAQKLDPSASBEVKSFE 361  
Db 237 F1NALQSGCNRWVKEIRKVTQLEBPSGSGISLOEMTFWNLBELLKIS-----OK 287  
Qy 362 KFGKRI-----LVKCNDSLFPNLQCCVAENEBEGPTTVEPEFVTLSPD-----IKYRKIS 412  
Db 288 RGEERTLTLEALKCG-----KRFHATVG--FDDNGLKQGLAVY 325  
Qy 413 ADFHVDLNFVSQMI--ATTSPALMNGSPETQSALRGI--LHEAM--QYRKQIGSVT 467  
Db 326 QVNTLMKEFPLSELVASATDVPKLMH-----AVVGIFELHRLKRLSTKYLQRA---- 373  
Qy 468 C8HPDILVARIKLEKLGOSITHCABEPMYSSDSKVAQKYLKAKACQRLQGYRPMFAM 527  
Db 374 -----LRLVEAISRDLSQLKLVSSY--NLKRTPIAE--FNEIMSQC----- 413  
Qy 528 AARTL.FKQASGMLDKVARFSAYIRODSNKLSDNMLK-----LADPR 570  
Db 414-----ALFSKDDDEYDK--FIALLR--DINKKGRDDBSKLNUKTAHVKLETRLMQLOLR 466  
Qy 571 KEKMAKT-----PVILGN-----LDI-----TIDNVSSDP 597  
Db 467 KOHEGRVTVIERLVKRGVNGSREBEQMLIDSSEKSPDEQVDAVEFLKAVDFLVDVSP 526  
Qy 598 NVVNSYITKQPER--CSKTPITEVEEVPFCIKHQPTIYV--NHLVYPRY--- 649  
Db 527 AMENA--FKRYEQOIGVETAITTRLSQLESNSNEMSPSIFERYALFPRIRIGA 582  
Qy 650 -----LKVD--SOKSFAKARNIAICIEFQSDDEDSQPKCIYGRPGGVFPR 695  
Db 583 IYEQTRLINRYKEDINELQARFTKRG-----EQYK--IMQVVGRLPPEA 627  
Qy 696 SAPPAVLHHQNPEDYDEIKIELPQOLMEKHLLLPFHVSQDNSSKSGSTKRDVETQV 755  
Db 628 KIMJIRNRYEROLORYKRVEDVLGQ-----WENHVDGQLADGDNFV 672  
Qy 756 GYSWPLKLDGRVYVTEQHIPIVSAN--LPSGYLGVELGKGRHYGPEIKVNDGSKL--IX 812  
Db 673 KLNTQPMFDEW--VESVQ--SQMWTLPNKILTYDVR-----QVGRMQLQK 715  
Qy 813 ISTHLVSTV--YTQDOHLH-----NPFQYCOQTESGAQALGENELVKYLLKSLHA-- 858  
Db 716 IYHSSSEVLKYKEVSHLKSMMGRVPLKYIVMAHQANQMPSATSL--IEAARTFASVNAAL 774  
Qy 859 --MEG--HVMIAFLPTILNQLFR-----VLTRAQEBVAVNVR 893  
Db 775 ASVQGVDSLASYKQIDIONOLIEGATLGWDSYKVDQYKLAETVTVYGERBELLNVVR 834  
Qy 894 V1-----IHVVAOCH-----EGLESLSRASY 915  
Db 835 IYNADNLVKSGRYDKETIENLITSIQGVDOQLSLQNYNLQAWNTLROJETTILARRV 894  
Qy 916 KYAYKAEPVASBYKTVHEELTK-----SWTILKPSADFLTNKYL 957  
Db 895 EDAIRKWTILVPSQSEV--BELERQVYVLPVKNVVVDLCKTACOTLYISSTR--ETREKYL 952  
Qy 958 K--YSWFPPDVLIKSMAQHLIENSQVLLRNQRF--PASVYHHAVETVNNMLPHITQKFD 1014  
Db 953 EOLYEHWSV-----CTAQWRISGKRQVMNNEIEBETVHN-----ILN--VMEGQACLEK 1002  
Qy 1015 NPEASGNANHSIAVFTKRCFTF-----MRCGFVKQKINNYISCFAPAGDKITLPEYFEBL 1069  
Db 1003 AYDCVNGIMSDEBYLSEWLSYOSLWVLAQOELFEMLGISLSKM-----KITMEIRKG-- 1056

Qy 1070 RVVCHNHEHYPLNLPMPGKRIQRYODLODYSLTDEFCRNHPLV--GLLIREVGTALQ 1127  
Db 1057 RLVPDQOTRKVIFPVSVBYGKAQ--QKILFRY--DYMHKEMLVKSGAV--VGDEMQ 1107  
Qy 1128 EPREVRLAISVKNLLIKHSPD-----DRYASRSHQ 1159  
Db 1108 KF-----FNSVSKRWVLEQSVGSGSTDTIGLISFVQSLKQTKSGDAVDLYRSSQRL 1163  
Qy 1160 ARIATLVLPLRGLLENNOR-----INRDSV-----PPVYNGMTVKOSLALPVN 1207  
Db 1164 LMQRYOPPAQWLYSENVGEMSAFTEILSLDASITQOMNLOTKYFAOEBELKRTVE 1223  
Qy 1208 PLVTPQKSTLDSNLIH--KDLGALISGIASPYTSTPININSYANA-----DSRSLISTDSG 1262  
Db 1224 TLTEMNKRPVGAQARPOBALNVITAFPAKUNKLLEENKORAKAVALDLSASAPBEG 1283  
Qy 1263 NSLPERNSEKS-----NSLDHQOSSLGNSVVRCDKLDOS-----EIKSL 1303  
Db 1284 DKLTVAATEELAMKQVMKALOPVYTGIDAEKKTWLS--VQPRKIRBSLBELMQLKOL 1340  
Qy 1304 --LMCFLYLKSMSDALFTYNNKASTSLMDFTISEVCLHQFQYMGKRYIARTGMHAR 1362  
Db 1341 PVKCRTYASYEHVKOMLHTYGRMMNLVAELSEALKERHMQ--MMKEMRYNNLSOLT 1397  
Qy 1363 LOQLGSLNLSLTFNHSYGHSDADVLHQ-----LLEANATEVCLTALD-----TL 1408  
Db 1398 LGQVW-----DADILRHETIKILLVAGGEALAEFLEREMRYQWNTVEV 1442  
Qy 1409 SLFTLAPXQUL-----LADGHNPLMKRVFVYVLCFLQKQISETALKONFTAL-- 1456  
Db 1443 ELVYNQKTRILKGMDDLFPNKLEHQNSLSAMKLSPPYKQFEESQSDERKLNKINAMFD 1502  
Qy 1457 -----RSLIYFPFTFEGRADMCALCYELIKCNSKLSIRFESOLLYFLMRNPD 1510  
Db 1503 WVIDQRRWVYL--EGLSSGSEISTLPEF--SSRPATITTDVIALM----- 1546  
Qy 1511 YTGKKSFVTHLQVIVSVOLADVYIGGT-----RQOQSLIINN- 1553  
Db 1547 --K------VAASPRLDVYVNMQAGORLLERLADMLAKIQXALGEBLERBSF 1593  
Qy 1554 --ANSRDLIKHTSFSSDVYDLTKRIRIYVLMATAQKHEHDEML-----VDL 1599  
Db 1594 PRFVFGDEDLLEIGNSKDIYRIQHLKMKPAGITADINEEDRSITAFHSREGEKYL 1653  
Qy 1600 QYSLAKSASTPELR--KTMUDSMARINHYKQNDLSAANCYVHVTVLVAEYLTRKAYQME 1658  
Db 1654 --VKIVSTDVAINDLQAL--BAEMKHTLARQLAASLTHFSKINIQTMTTDDYVW- 1706  
Qy 1659 PRLPHSHSACLRBSRGVFRQGTAFRVTIPNI--DE--EASNMEDVGMODVHFNEDVL 1714  
Db 1707 -----LDKRPAYV-----TLTAEIWMCEMETITLADGGAENV--EQAV 1744  
Qy 1715 MELLEQCADGLWKAERYELIADIKYLIPIYKRRDPERLAHLVDTLHRAVSKTEVNH- 1773  
Db 1745 VKTLEBLADSVLKQEP-----PI--RRKQMBAL-----ITELVAK 1777  
Qy 1774 --SGBRLIGYFRVA--FRG--QAAQOFTDBETV-----EGPREBBDGSEYIYK 1818  
Db 1778 RDTCKLVSMMKIRANDFGMLQCMFYPDPKQVDPVRCVVYKMANSOFFY--GEYIGI 1834  
Qy 1819 EPKL--TPLS------ISQRL--LKYSDKFG-----SENYMIDSGKVNPKDLSKYA 1864  
Db 1835 QERLVRTELDRCYLTMTQALHSRSGPFPBAGTKTSYKAL----- 1878  
Qy 1865 Y1QVTHVLPFDEKELQERKTEPERSHNIIRPMF-----EMPTQYSGQGVBEQCR 1918  
Db 1879 -----GHQLGFRVLVFNCDETFDFQAGRIIVGL--CO- 1909  
Qy 1919 RTILTAHCFPVYKRIIVMTQHTDLNPIR-----VAIDENSKYVALEROLCSSAEVUMI 1974  
Db 1910 --VGAMGCF-----DEFNRLBERMLSAVSQOIQTIQAVRAVAGGDSVDLV 1952

OY 1975 KLOJLQGVSVQVGNAGPLAVARAFLDITTKRYPPNKKVLLKEVER 2021  
 DB 1953 GKRLLVNSNIGIFITMNPVSGRSNL-----PDM-----LKQLEP 1987  
 RESULT 8  
 RBP1\_PLAVB STANDARD; PRT; 2869 AA.  
 AC 000758;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 1 precursor.  
 GN RBP1.  
 OS Plasmodium vivax (strain Belen).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 NC NCBI\_TaxID=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92315338; PubMed=1617731;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 RT merozoites";  
 RL Cell 69:1213-1226(1992).  
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M88097; AAA29743.1; -  
 KW Malaria; Receptor; Signal; Transmembrane.  
 FT SIGNAL 1 17  
 FT CHAIN 18 2869  
 FT DOMAIN 18 2807  
 FT TRANSMEM 2808 2826  
 FT DOMAIN 2827 2869  
 FT SITE 1030 1032  
 FT SITE 2599 2601  
 SO SEQUENCE 2869 AA; 330213 MW; B9DBE4422058BCFF CRC64;  
 Query Match 1.7%; Score 187.5; DB 1; Length 2869;  
 Best Local Similarity 17.1%; Pred. No. 0.028;  
 Matches 433; Conservative 408; Mismatches 885; Indels 809; Gaps 118;  
 OY 34 SEVVRGSVLLAKKLEPLDYENVIVOKTQIANDC--LRMLLPYDDPQTILARQGR 91  
 DB 480 SKVTKSNELLS--TIIDLGKSAATVAQESTPDEBCNKIKTEAEKVKDAEDICEKNEDI 537  
 OY 92 YICSTYPAKAE--EAGSLPTECICKY-----NSDWHLVNYK-----YE 129  
 DB 538 YV--ELPSEDEITIDKINDLODLT--DQMKYKDIIVNNSEFISIKRYIENLKETEY 593  
 OY 130 DVSGEFRQLPNKVVKLDPVHYVEVDEVDKEDASLSGOKGKITKGWLYKGNMS- 188  
 DB 594 TELNDIGIKLENDTSKYNFYLMQIRIKINTEKTKIDESLQIVTEK-----FYKEILDSEK 644  
 OY 189 -AISVTRSKRFRFFHILQIGDSYKFFELKDLQKPKSIFLGLPLGVSVFRNNKVRRA 247  
 DB 645 EKYTELKIEFKSVTEINRLQDQ--ESAPDLHEQIKEL----- 682  
 OY 248 FELKMDKSSYL-----LAADSEVEMEMTTLNKLQNLFE--AAQOEKRGNDSH----- 296  
 DB 683 --DKMAKRYHLLKELSLKSKSSVYFTENNELNLTASYDNMBGFSKAKKADADINALYN 740  
 OY 297 -----EDDEQSKLEGGSGGL-----DSYLPRLAKSARB--AEIKLKESRVLKFLY 339

DB 741 SVYREDINALIBEVEKFTVENKSTLEMLKDEMEEKLQDAKETPAKLPVSDDKLTDVY 800  
 OY 340 LDPDAQ-----KLDSSAPPEVYKSEEFKGRILVKCDL--SFNIQ--CCVAEN 385  
 DB 801 TKMSAEVTTNAEGIKKEIAQKQFENVHKMKKESDAPSTKFEALQNSMOQYQDEGDAIEGH 860  
 OY 386 EGGPTTNVPEFVTLSTL--FDIKYNRKISADPFVLDNHSFSVROMIATSPALNAGSGPETQ 444  
 DB 861 KQNSBEKEBEYFKMSVEEDLSRETEBEQYTKHKNNSRRK--GEISAEIN----- 911  
 OY 445 SALRGILHEAANQYKQGI--FSVTCRPHDIFLVARI--EKVLQGSITHCAPY----- 494  
 DB 912 --MREVINIKESQLNVYGVIEKYFSLIGDQNEVSTAKAKKIVDSLRDKIDQYETEERK 969  
 OY 495 -----MSSDSKYAQAQVLYKAK-----QACQQLGQYRMP- 524  
 DB 970 EKTSAVENTVSTIQLSKAIDSLKRLNGSINNCKRYNTDILRSKIKITLREEVGKEMPK 1029  
 OY 525 -----FAMAAATLFPKASGNDKNAFSAIYRQDSNKLNDMLKTLADFR----- 570  
 DB 1030 RQDKCGENTTALLKSL--RDKMGKINELKNDORLMSLDTK--EDLKFYSSESCKIHL 1085  
 OY 571 --KPEKAKLPV-----ILGNLDTIDNVSSDPN-----YVNS--SYLEPT 608  
 DB 1086 SKDQKGPQDPLNRIIDEMEDIKRDVDELNVNVQVISENKVTLFKNNSVYIEAMHSHINTV 1145  
 OY 609 QPE--TCSKTPITFEVE----- 1145  
 DB 1146 AHGITSNKNELIKSVKEVEDKLNVEQNEQYKVPENKQLEAIRGSMKKEVINHG 1205  
 OY 633 TQPYTYTNHLYVVPKYK-----YDSQSFAPKARNAICIEFK 671  
 DB 1206 VSEMTQBSTANTLKSNAKKNENHDLBELNKTQGMQDIYKLLKIBELKEGVNLEK 1265  
 OY 672 DSDEE--DSQPLKCIYG-----RPGCVFTRSAP----- 698  
 DB 1266 DANERANKKPEPERENIIGHVLERITVEKDKAGKVEEMNSLKTIEKILQETSDSQNE 1325  
 OY 699 ---AAVLHHQNPPEYDEL--KIELPQOLHEKHLILTFPHV-----SCDSN 740  
 DB 1326 LVTTITTHLENAKYEDVIRKNEEDSTQLEBKAKSLTLEBMKKLVQOVNNNLSAIOG 1385  
 OY 741 SKGSTKK--RDVVEQVGSWMLPLK-----DGRVVTSE--QHIPVSA 779  
 DB 1386 NAGISKEINELKGVIELISTVYSILRYVKNSSSEVSFQSLANGERTKAGEEKNSA 1445  
 OY 780 NLPSEY-----LGVOELGMGRHYGEIRKWDGK--PLIKISTHLVSTVYTQDQHL 829  
 DB 1446 RLAEAEKLEKQIVKDLVSDI-----DDKYKIEGIRREILKMK-----ESALTFWESE 1495  
 OY 830 NFOYCKK-----TESGAQLGNELVYKILMSLHMEGV 863  
 DB 1496 KFKQKSSHMEKAKKKEIKVLEKKNNGDGKANTIDSOHEEYGNVSAEHAHFTVEQV 1555  
 OY 864 MIAFLPTILNQLFRVLTQTOEVAVNTR-----VIHVVQACHEEGLESILRSYV 915  
 DB 1556 -----DKTAFCEISIVAYTTQNDLNPENSLMEKVKCKCKNDAP----- 1596  
 OY 916 KYAYTAEPY-----VASEKYVHEELTKSMTTI--LKPSAD 949  
 DB 1597 KYSATLKPYDGRKIRAVSENERKISLEKAKAVEKSSQNDVSTKSLQIDNCRQQLD 1656  
 OY 950 FLTSN--KLKYSWFFPVULIKSMAQH-----IENS--YVKILNORFASVYHNAVET 999  
 DB 1657 SVLSNIGVQVAAQVFDSPADSKMSKSVLPISLGAELKSLDYKAKA-----SEKNET 1711  
 OY 1000 VNNMLM-----PHITQKFRDNPESKNNANSLAV 1028  
 DB 1712 VQNEKSRINVEGSLTIDKITDLENDLLKMKQYEGGLQIKENADKRR--SNFEL-- 1768  
 OY 1029 FIKRCFTMDRGVFEKQIINNYSICFAPDPT--LFEYKFEFLRVVCHENY--IPLN--LP 1084

```

Db 1769 -----VSGEINALL-----DPSISIFIKLKLKEDMTGDKMNGVKNNEIH 1809
Qy 1085 MPFGK--GRIGRYQDQLDYSLTDEPCRNHFLVGLLRVWTAJOE---REVRLI----- 1135
Db 1810 GEFYTSYNLIERHLSNATDYSTFEKAOS-----LRRLAEKEEHLARRREEAIFLIN 1862
Qy 1136 -----AISVLKNLKHSPDDRYSRSHQARIATLYLPLFGLLINVRINRVDSPPP 1189
Db 1863 DIKVEYSLKLKEMKVSAYEGMKRDHTS-----VSLQVDMKTI----- 1904
Qy 1190 VNAAGMTVQESLALPVPNLPVPOKSTIDNSLHKD-----LGAISGISPYTTSTP 1242
Db 1905 VBLKTLNDISECCSVLNNVSVIVK--KVESHADVRDANSWMSYTLNAPLSDBA 1962
Qy 1243 NINSVA--NADSRGSLSTD-----SGNSLPERNSKSN--LDHQOSSTGNGSV 1289
Db 1963 KISSGMEFPAEMK-SNFKTDLBELFISVINSNELKLKEQDSNDVIOKERESEOLAK-- 2019
Qy 1290 VACDKLDGIEIKSLMCPLYILKMSMDALFTYWNKASTEELMDFTTISEVCLHOPQYNG 1349
Db 2020 -----DATDIYVNIK-----LKNFENEKLEBAKN-----BEVSEKVRBAL 2056
Qy 1350 KVIYATATG-----MHARL---OQUGSDNSLTENHSGHSDADVLHOSLEBANATEV 1400
Db 2057 KRLSQVEGRCHRENFHRLDTELEENLKQVVTIYRDKISERESGLQEMENEMTYSN- 2115
Qy 1401 CTALDPLSLFTLAFKQOLLADHGNPLMKKVDVYLCPLOKHQOSETALKNV---FTALR 1457
Db 2116 SITOLEGIYVSAGESKEDI-----EKL-----ERSHEEMENISEKSTID 2155
Qy 1458 SLIYKPSSTFEY-----GRADMCALCYEILKCNSSKLSIRTE---ASQLLYELMN 1507
Db 2156 SKYIEMNSTIDELYKGNCOQAMWISLSTANMKTSSKLMINKENKTEKVDVYIKON 2215
Qy 1508 NNDYTG---KSGFVTHLOVISVQL--ADVAGIGGRRO-QSLSTIINN- 1553
Db 2216 SSSDTGVEYELKGFYSKL-TRESSASEIYQMDVTSVNFAPKEKESLNAIRIDKELULF 2274
Qy 1554 -ANSD-----RLIKHTSFSSDVOKLTKRIRITVLMATQOMKEHNDP--EMLVDLQY 1601
Db 2275 HONSDSIYEGVQNMALYDKLNEKREMDLYRISETKYLKOMESHSTDFVEKPEMLK 2334
Qy 1602 SLAKSYASTPELRTKTLWDSMARLHVXQGLSEAMCYVHTALVAEYLTRKEAVOEPPL 1661
Db 2335 GNNETN-----NKSLEKEKKLKSVND-----HMHSSEAMMI---KQGLKATPSS 2376
Qy 1662 LSHSHSACLRBSRGVFRQGTAFRYITPNIDBEASMEVDGQOVHNEVDLMELLBOC 1721
Db 2377 VONINN-----IYGVIEAEVKTLEID-RDYGDMYOIVIEHKQO- 2414
Qy 1722 ADGLMKAIHVELIADIYKLI--IPYERKRDEPERLAH.YDTLHRA---YSKVTEVMHSG 1775
Db 2415 -----FSLIDBTNALMDIDIEFKKENNINLEAVTETIHRVNDYIEKTLNKLVOA 2465
Qy 1776 RBLLGTYFVAFPGAAYOFTDSESTDVEGFEDDEKELYIY-----XE 1819
Db 2466 K-----TEBOILENIKQNDMLNIFLKVYSIIEFENYVKKKE 2505
Qy 1820 PKLTPLEISORLLKYSDKFGSENVKMIQDSGKVPKULDSKVIATYQVTHVIPFDEKE 1879
Db 2506 SILNDIYE-QERLLKI-----GEHL-----DEIKRNVITFLSSEYEDQKEMM- 2550
Qy 1880 LQERKTEPERSHNIRFEMFPTQGRKGQGVBOCKRITLTAHCPYVKKR---IP 1936
Db 2551 LBEKSKSMNNTYSI--YELERANEENNRDAKQIKD--DTIINSV-LEAALQKRDMDA 2604
Qy 1937 VNYQHTDLPLEVALIDEMSKVAELROLCSSAEVMDIKLOLQGSVSVQVNAAGPLAYA 1996
Db 2605 IFSQMSADBNPNEYKSAE--KWNENANEIIROLEVGLRIGQVQDSESI----- 2652
Qy 1997 RAFPDDTNKRYIPDNKVKLKEVFRQVPAAGQALAVNRELKEDOLEVQOEMKANYRBM 2056
Db 2653 ---LSEWNSKKAIEK-----EKTAALATISNNRE-----EERBARVQEM 2691

```

```

Qy 2057 AKE-----LSEIMHEQ 2067
Db 2692 SMNNDPTQSETHSE 2706

RESULT 9
HMW2 MYCPN
ID HMW2 MYCPN STANDARD; PRT; 1818 AA.
AC P75471;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome high molecular weight protein 2 (Cytochrome accessory
DE protein 2).
GN HMW2 OR MP310 OR MP526.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97252497; PubMed=9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Hermann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT pneumoniae cytoskeletal protein HMW2 and cytochrome."
RL J. Bacteriol. 179:2668-2677 (1997).
CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC -----
CC EMBL; AE000051; AAB96174.1; -.
CC DR EMBL; U59896; AAB52527.1; -.
CC DR PhosSite; P75471; -.
CC KW Cytochrome; Structural protein; Coiled coil; Complete proteome.
CC FT DOMAIN 31 880 COILED COIL (POTENTIAL).
CC FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
CC FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
CC FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0CFBC0 CRC64;

Query Match 1.7%; Score 186; DB 1; Length 1818;
Best Local Similarity 17.5%; Pred. No. 0.017; Indels 710; Gaps 88;
Matches 369; Conservative 310; Mismatches 716;
Qy 214 FEFLDLOKEPKGSIFLGLVGSFRANKVRFAPELK--MODKSYL-----LAA 262
Db 95 PEFQGNVWD-----RVALNNKARFPODELRLQENNAFNSRYAMADPQS 143
Qy 263 DSEVEMEMTITLNF-----ILQNFPAAMQ-----KRGDSH3DDQSLGSSGSL 311
Db 144 NYQLKLDQFOALIDQONOTIKOLNIOANOGLLDONVORLQONHSLDOQR----- 195

```



QY 312 DSYLPELAKSAREAEIKLKSESSEKVLFPYLPDAQKLDSSSABPEVKSFEKFKRLIVKC 371  
D 196 DALVEYEDHLYNEL---YELNQKRLVIGIEVATYQDVLVSADAELOVNETTIO----- 246  
QY 372 NDSFNLQCCVAENEBGPTTNVEPFVYTLSPDIKYNRKISADPHVDLNFHVSQMIAT 431  
D 247 NQANFKQOC-----DAYMAOLK-----QVEQOIQT 272  
QY 432 SPALMNGSGPETOSALRGILHEAMQVPRKQISVTCPPHPIFLVARIKVLGSGITHCA 491  
D 273 KQSLV-----DEESTLKVRLNDA-----DFYINSRLAE-LDDLTSKIN 309  
QY 492 EPMKSSDSKVAKVAKKQAQCORLQGYRMPFAMAARTLFFDASGNLDKNARFSAIYR 551  
D 310 ERDFVSEKQADYKASLANLTKERERLSAEKSFBRRLNTALDINRMEQENALFKHLE 369  
QY 552 Q---DSNKLNDMLKILADFR-----KPEKAKLPVIL-----GNLDTI 589  
D 370 QOQYEFERKQOESLKLKLETHKOLOKRIGEFKIESEKSEALLIORRELEKRRRIDDL 429  
QY 590 DNVSQD-----FPRVYNS-----YIPTKO-----FENC 613  
D 430 TQMSLEYEQORINOVULKEKGRVOQHFOVLVHAKKLDQKRYLABQKRIDEEQIFKL 488  
QY 614 SKPTIFEEVEEFVPCIRPKTQPYTIYNHLYVPKYLK-----YDSQK 656  
D 489 -KEKITERRELEKLYLVKKQODQKENDLIFEKOLROYQADFENBIEKQONELFASQK 547  
QY 657 SPKAKNIAICIFKQSDSEDSOPKCIYGRPGGVPFTTSAPAAV---LHHNQPFYDE 713  
D 548 SLOKS-----FTQKKKKEALNOKAKI-----AEDMAHLKQNKHHADEIF-- 590  
QY 714 IKLELPTQHEKHLHLLTFPHVSCDNSKSTKRDVETQVGYSMPLL-----KQGR 767  
D 591 LBESEPHNLOEKKIL-----EARTQPDNRVS-----LISAFKQKQAB 629  
QY 768 VTTSEOHIPVANSIPSGYLGQYELGMRHYGPRIKAVDGGKPLKISTHVLSTVYTOQOH 827  
D 630 LVKQKQSL---EQLTAFNKEQE-----AVERDKMDRLANLEKQKEMLGDKVHQFDEN 679  
QY 828 LNHFPQOCTEGGAOLGNELVKYLSLHAMEGHVMIAPLPTILNOLFRVLTFRATQOEY 887  
D 680 SLNISKKLARELAIFKEKELEMAQKQSLDNNN---NAGLKLQDLRLSESL-KTERLEI 736  
QY 888 AVNVTREVI-----IHVVAQCHEEGESHLRSYVKTAVYABPYVAS---EYKTAHELT 937  
D 737 EASKEHILDPYDESSSRIRAD-YESDLOARL-AEVKTLKNOQETAKSERELKVALEKDN 794  
QY 938 KSMITLIKPSADELTJNKLLKYSWFFPDVLIKSMAOHLIENSRYKLLRNQRPASVYHAAV 997  
D 795 QAKKAFILQ-----IRKQOULE-----IASVQOQLAQ-KANLLKNOQ-AELDQKT 836  
QY 998 EYTVNMMLPHITQKFPDNPEASKANANSLAVFIKRCFTFMQDGFVKQIINNYSQAPGD 1057  
D 837 EETLEAFLBEDDTDK-----KELEKALHS-VKSKQELERERSFLQKQREFA----- 882  
QY 1058 PKTLFYEKFEFLVVCNHEHYIPLNLPMPFGK-----GRTOR-----QODLOLDVSLND 1106  
D 883 -----EHVAGFKQV---HFKTTQMORLSEFNKQOQSEQIKRETELKIAFADLKQDOLP- 934  
QY 1107 EFCRNHFLVGLLREVGTALQEFFREVRLTASVLKNLLIKHSFDDRYASRSHQARIATLY 1166  
D 935 ELQIKN-----QEFQOIEQKHELELLA---QKQABLKQBLEQATYALASQ----- 976  
QY 1167 LPLFGLLIENVQRIINRVDSPPVNAQMTYKDESLALPANNPLVTYQKSGTLDNSLHKL 1226  
D 977 -----DQD-----TYQAKILDLARQOHE-- 993  
QY 1227 LGALISGIASPYTTSTENINSVRNADSRGSLISTDSGNSLPERNSEKNSLIDKHQOSTIG 1286  
D 994 -----LELRQNAFQASISLKLQREQLTNQVKVLHGEIKGRHEKLTLLK 1036

QY 1287 NSNVACDKLDQSEIKSLMLCFLYIILKMSDDALFTYWNKASTSELMDEFTTSEVCLHQO 1346  
D 1037 DDLAEKEKQDQK-----KQABINQRFQFQENEVAD-----FD 1069  
QY 1347 YMGKRIYATGMHARLQOLGSLDNLSTFNHSGSDADVLHQSLLEANIATEVCLTALD 1406  
D 1070 QAKKR-----ELOELNQRIRNLB-----OSNMSLAKK----- 1096  
QY 1407 TSLFTLAFKQOLLDHGNPMLKKVFPDYVLCFLQKHQSETLKQVFTALRSLIYKFPST 1466  
D 1097 -----RNQTLTDEFA---LKRKY-----QHTQTRVQVLNTQIKFLELKKN- 1134  
QY 1467 FYEGRADMCA---ALCYELIKCNSKLSIRTE-ASOLLFYLRNNFDYTGKSEFVRTHL 1522  
D 1135 -FOKASDEBALQKALLIKRLRSFASKLQGFQBALAIQKLEFQKR---DEQKSEINNAKL 1190  
QY 1523 QYIISVSQILADVIGIGTRFQOOSIINNCANSQ---RLIHGTFSSDVKDLTKIRIV 1579  
D 1191 Q---LEQFKLEKQNFDEAKQKQLEFQDQCORLQVERKRLK----- 1228  
QY 1580 LMAAQMKHEMDPEMLVDLOYSLAKSYASTPELAKTMLDSMARLHVKNQGLSEAMCYV 1639  
D 1229 -----OKLVOLK-NLSKSYLT-----YKRRADLSQOOLQHK 1258  
QY 1640 HVTALVAEYLTRKEAVQWEPPLLPKSHSACLRSRGV--FRQGTAFRVTIPNIDEAS 1697  
D 1259 Y-----ANLLEBKELQTKRALDKGRALYQKMAQFSELQKQKOLLSAQKQVDDKR 1313  
QY 1698 YMEDVQMDVHNEDVLMELLEQCADGLMKARVYLADIYTLIPIYEKRDPERLANL 1757  
D 1314 LLEQ-----NORHLONLSET-----KKRQSLHDHINK---FQRRK----- 1348  
QY 1758 YDTHRAVSKYTEVHNSGRLIGTVFRVAFQGAQOQYFTDESTVEGF-----F 1807  
D 1349 -----EAVSSIANSHKL-----KQEGSELQIGLOKLSLKTQI 1382  
QY 1808 EDEDEGEYIYKRP---KLTPLSEISQRLLYSDKFGSENVMIQDSKVPKDLDSKYA 1864  
D 1383 EQFESKLYQQRREKLDQRRTLSKL-HRELKAQNEATYAHKNREVL----- 1426  
QY 1865 YIQVTHVIPFPEBEKLOEKRTPEFSNINIRMPFE-----MPTQOTOK 1907  
D 1427 -----IENYKKELORLTTEKSEFDDNNKRLFEYFRKIRINEIEKKAHKTVALEETOK 1479  
QY 1908 ROGVEEOC-----KRRITLFIHCFPYVKRIRPWOQH---HTDLNPI----- 1948  
D 1480 KHLVETEVLKHLQKOSTISKGQELKEIKEVASDISHTKQRELSLSLHQKLLQKN 1539  
QY 1949 ---EVAIDE---MSKRVABELQCSSAEVDMIKLOLQOG-----SVS 1985  
D 1540 LAEREREINNKSLLTQIKOTAKQKLSKEKARILKLEKRAVEQOYQAEITRLKTRNAD 1599  
QY 1986 VQVNAQPLAVARAFLDDNTKRYP-----DNKVLLKEVRFQVE----- 2025  
D 1600 LEKNKNKHLFPLFKINGIDNMVPPYPFWPQOQOEDSSNQIRHLFQOOLQFMOQRYEN 1659  
QY 2026 ---ACGOALAVNERLIKEDOLEYOSEMKAN---YREMAKESLIMHBOICPLEKTSYVL 2078  
D 1660 ELTELRRQBALIEKLDQIOLESQLSAKKNDEKVEQMMQKLEETBEQKLSAFQKINAL 1719  
QY 2079 PNSLH 2083  
D 1720 AEOIN 1724  
RESULT 10  
CENE HUMAN  
ID CENE HUMAN STANDARD; PRT; 2663 AA.  
AC 002324;  
DT 01-JUL-1993 (Rel. 26. Created)  
DT 01-JUL-1993 (Rel. 26. Last sequence update)  
DT 30-MAY-2000 (Rel. 39. Last annotation update)  
DE Centromeric protein B (CENP-B protein).



GN CENPE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93024922; PubMed=1406971;  
 RA Yen T.J., Li G., Schar B.T., Szilak I., Cleveland D.W.;  
 RT "CENP-E is a putative kinetochore motor that accumulates just before  
 RT mitosis."; Nature 359:536-539 (1992).  
 RL Nature 359:536-539 (1992).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=95196755; PubMed=7889940;  
 RA Thowar D.A., Jordan M.A., Schar B.T., Yen T.J., Wilson L.;  
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
 RT microtubule motor."; EMBO J. 14:918-926 (1995).  
 RL EMBO J. 14:918-926 (1995).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98437347; PubMed=9763420;  
 RA Chan G.K.T., Schar B.T., Yen T.J.;  
 RT "Characterization of the kinetochore binding domain of CENP-E reveals  
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";  
 RL J. Cell Biol. 143:49-63 (1998).  
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
 CC OF THE KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
 CC AND/OR SPINDLE ELONGATION.  
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
 CC CONSENSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTHIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z15005; CAA78727.1; -  
 DR PIR: S28261; S28261.  
 DR HSP: P17119; 3KAR.  
 DR GeneW: HGNC:1856; CENPE.  
 DR MIM: 117143; -  
 DR InterPro: IPR001752; kinesin\_motor.  
 DR Pfam: PF00225; kinesin\_1.  
 DR PRINTS: PR00360; KINESINHEAVY.  
 DR SMART: SM00129; KISC; 1.  
 DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.  
 DR PROSITE: PS0067; KINESIN MOTOR DOMAIN2; 1.  
 DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
 KW Cell cycle; Centromere.  
 FT DOMAIN 1 335 KINESIN-MOTOR.  
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
 FT NP\_BIND 86 93 ATP (BY SIMILARITY).  
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8B8 CRC64;

Query Match 1.7%; Score 186; DB 1; Length 2663;  
 Best Local Similarity 17.1%; Pred. No. 0.031;  
 Matches 441; Conservative 380; Mismatches 949; Indels 814; Gaps 112;

QY 19 ALSKPTAELROS--VSEVRSGLAPKLIPEPD---YENVIVQKTOILNDCLRE 72  
 DB 319 ALGFSTAKYMKNTPVNVEVSTDEALLKRYRK-ELMDLKQLEVELEFRADAMEMDOLA 377  
 QY 73 MLFPYD---DPTALIRROGRYICSTVPAKAEERAS--LFTVECTITVSDMHLVNY 126

DB 378 OLLEEKDLLQKVONEKIENTRLMTLTSSSLTLOQELKAKRRRYTWGCIKGIN-----KM 431  
 QY 127 KREDYSGEPROLPNKVVLDKLPVHVY-EVDEYVDKDDAASIGSKG3ITKMGW----- 180  
 DB 432 KNSNTADQNPNTITTTKHLSTVLAREIBSVCSSESDVS--NTLDTLSIENNPATK 489  
 QY 181 -LYKGNMNSAISVTNRSFKRRPFHLIQLGDSYKEEF-LKD-----LOKEPKG9IF 229  
 DB 490 LNNQINSELSNLSADYDNLVLDYEQRTKEEMELKLEKNDLDBFEALERKTKQOE 549  
 QY 230 LGFLYGVFPNNKVRFPFELKMDQKSSYLLAADSEVM---EEMITLANKIL-QUNF 283  
 DB 550 MQLIHESNLKXVLGH--REYVNDLENEL--SSKVELREKEQOIKQLQEYIDQKLE 604  
 QY 284 EAAWKEKNGDSHEDEOSK---LEGSGSGDSYLPFLAKSAREAEIKLK----- 330  
 DB 605 NIKMDLSLSIESIEDPKMKQTLFPAEYVALDARKESAFLSBNLEKKEKELATTYQ 664  
 QY 331 SESRYKFLYLDPAQKLDSSAEPVKS-FEE--KFGRIILVKC-NDLSPNLOCCVANE 386  
 DB 665 MENDIQLYOSQLEAKKKQVLDKELOSAFNEITLTSLIDQKVPKDLLCNLEL----- 718  
 QY 387 EGPITNVEPFTYTLSPDIKYNRKISADPHVDLHNFVRQMIATTSPPALANGSGEPTOSA 446  
 DB 719 ECKITDLQK-----ELNKEVE-----ENBALREBYITLSEL---KSLPSEYER 758  
 QY 447 LRGLIHEAMQYPKGIFSVTCPHDIFL-VARIEKVLQGSITHCAPYMKSSDSKVAQ 505  
 DB 759 LR-----KEIQKSEELHITSEKDKLPSEVNHKSRVQGLI---EETGKQDLATQ 809  
 QY 506 KVLKNAQACORLQGYRPFMAATLTFKDSG-----NLDKNA-RFSATY----- 550  
 DB 810 SNYKSTQEPQNFKTLHNDPEQKYGVALEENBRNQELIVNLSKKAQKSDSLGALKTELS 869  
 QY 551 -----RODSKLSNDMLKULADFRK-----PERNAK----- 577  
 DB 870 YKTQELQKTEBVOGRLEMEQLKEQLENRQSPLOTVERBKTLLTEKLOQLTEBYKTLQ 929  
 QY 578 ----LPLVIGNDITDNVSSDFPNVNSYIPTQF-----ETCS--KTPIYF 620  
 DB 930 EKDDLKQLOESQTERDQKSDIHDTVMNNTIDTQELNNALESKQIHQETINTLSKISE 989  
 QY 621 EYVEFVPCIPKHTQPYTYTNHLYVVPYKLYDSQKSPAKARNIAICIEPKSDSEDSOP 680  
 DB 990 EYVSRNL-----HMEENTGETKQEF-QQKMGVGDKKQDLBAKNTQTLTADVNDMLIEQO- 1042  
 QY 681 LKCIYGRPGGVFTRSAPAVLHHQNPFEYDEI---KIELPTOL----- 722  
 DB 1043 -----RKIFSLIQEKNELOQMLESVIAEKQKTDLKENIEMTIENTQEBELR 1088  
 QY 723 -----HEKHLILTFPHVS--CDNSKSGSK-----KR 748  
 DB 1089 LIGDELKKQOEIVQAEKNHAIKKEGELSRCTDLRAVEBEKKEKSOQLQEKQOOLLNVOE 1148  
 QY 749 DVEETQVGSWMLPLKD---GRVVTSE---QHIFVSANLPSGYQYQELAGRGHYGPEI 801  
 DB 1149 EMSQWQKINELNKNLKNLKEKLELHMERBERLELAQKLENTY-----BEV 1195  
 QY 802 KWDGSGKPLKISTHLVSTVTTQDQHLNFPQYCOKTESGAQALGNELVKYIKSLHAMEG 861  
 DB 1196 KSITEREYVLK---ELQKSPETERDHLRGYIREIEAT-----GLQTKREIKI-----A 1240  
 QY 862 HVMIAFLPTIINQLEFVLTAT-----QEEVAVNVTVIIHVNAQCHEE 905  
 DB 1241 HHLKEHOETIDELRVSSEKTAQIINTQLEKSHTKLOEBIPV-----LHE- 1287  
 QY 906 GLESHLRSGYKAAVSEPYVAEYKVAEELTKSTATILKSSADFLTNSKLYKSWPFED 965  
 DB 1288 ---EGELNPNVKVSTQETNNEBELTQSTTOSTTLARLMEBELRLNKRQOE-----Q 1341  
 QY 966 VLKSMAGHLIENSKVKLLRNQRPASVYHAHVETVNNMLMPTITQKFPDNPASQVANH 1025  
 DB 1342 BEIKSLTR---ERDMLKTIK-----EALVAVHDLKEHIRETLAKIOESQSKOROS 1389



FT PEPTIDE 607 737 E (POTENTIAL).  
FT PEPTIDE 738 869 F (POTENTIAL).  
FT PEPTIDE 870 1003 G (POTENTIAL).  
FT PEPTIDE 1004 1074 H (POTENTIAL).  
FT PEPTIDE 1075 1208 I (POTENTIAL).  
FT PEPTIDE 1209 1342 J (POTENTIAL).  
FT PEPTIDE 1343 1475 K (POTENTIAL).  
FT PEPTIDE 1476 1545 L (POTENTIAL).  
FT PEPTIDE 1546 1557 C-TERMINAL EXTENSION (POTENTIAL).  
FT CARBOHYD 997 997 N-LINKED (GLUCNAC. .) (POTENTIAL).  
FT VARIANT 946 946 H -> Y.  
FT VARIANT 948 948 L -> V.  
FT VARIANT 955 955 L -> D.  
FT VARIANT 974 974 E -> G.  
FT VARIANT 990 990 G -> A.  
FT VARIANT 1543 1543 A -> V.  
SQ SEQUENCE 1557 AA; 182600 MW; 33717911CF55BD4B CRC64;

Query Match 1.7%; Score 185; DB 1; Length 1557;  
Best Local Similarity 16.2%; Pred. No. 0.015;

Matches 269; Conservative 262; Mismatches 582; Indels 546; Gaps 69;

QY 490 CAEPYKSSDSSKVAO-----KVLAKAKACQRLGQYRMPAMATL-----FKDA 536  
DB 214 CKKYFMQIANSDEVEKIKSLNDELIRHVYKNVAVARLNGELKEFAVVMETLCEVDLAFKAR 273  
QY 537 SGNLDKNAFSAIYRQDSNKLSDMDLKLADPR-KREKVAKLPIVLGSLDITIDVSSD 595  
DB 274 KNDIDDKINRLISMNTDEOK--QVVKYADGRSADIFAKIPEFLSSIDGAGYAAK- 329  
QY 596 PNNVNSVSIPTKOPETCSKPTTFE-----VEEFVPCIPKHTOPTTYTNHL 643  
DB 330 -AQIOKECYKMMVEVATAEIALHEIHEDHGCGRKVFETIGRLPE----- 376  
QY 644 YVYPKLTKDSQKSPAKANIALC--TEKSDDEBSQPKCYGRGGRVPTTRSAFAV 701  
DB 377 -----DRKLEVEKDLPECEKIMWYRDHGDHNSHK-----HGA 407  
QY 702 LHHNHN-----PEFYDEIKELPTOLHEGHHLLLPFHVSCDSSKSGTKRD 749  
DB 408 HHHHHHLAVRRRHLYAIKFLDMLK--PEQGHLEKI-----SNS-GAIFDDV 453  
QY 750 VVETOVGSMPLRLKXGRVVTSEONIPVSAHLPSGYLGVOELGMRHRYPEIKWVDGKP 809  
DB 454 IAEVKKFYGLP-----BEKIELAKAKFKSQCYDW----- 483  
QY 810 LKISTHLVSTYTTODQHLHNFQYQCKTESGAQALGNELVKYLKSLHMEGHVMIATLP 869  
DB 484 -----VKEVATSEEM-----NDIMK-----MHESKNH-----S 506  
QY 870 TILNQLFRLTRATQSEVAVNVTIRVLIHYVAC-----HEGLGSHLSRYVXK 917  
DB 507 DLMKRLTELENRLTEQO-----KHTLEHREVCLGMEVQNTNKKQKQSLSEAMDYLSW 561  
QY 918 AYKAEF-----YVASEYKTVHELTSTMTLLKPSADELTSNKLKLYMFPFDVLKSM 971  
DB 562 MDEDEKXKAIYETSNRQTFYDELK-----IMESSEDEVAKATK-----LEAAKCY 612  
QY 972 AOHLIENSVKLLRNORFPASVYHAAVETVNMIMPHITQKFRDNPASKANSHLAVFIK 1031  
DB 613 GTNIIIGEEVNDIIRKMKKNGATFEESISNREVDLEIGTIDSDR-----KEKAYRMSKCK 666  
QY 1032 RCFETMDRGFVKQINNYISCFARQPKTLFEKFE-----FLRVVCH 1075  
DB 667 KLYSL-----GHSKQLOQYDFENVLQKYLTMWLDSSQKMLRTMSDN 707  
QY 1076 EHYIPLNLMPPF-----GKGRIGRYOQLQDYSLTDFECRHHFLVGLLREVGTALOEFE 1131  
DB 708 KKKIYKKIIDYFDGTIGEVKEKAVELO-----ACNHYIKSVGEKAMEIKQKLE 759  
QY 1132 VRLIAISVLTNTL-LIKHSFDDRYASRSHOARIATLYLFGILLNVORIN----- 1181  
DB 760 BKSSSEIYAKVEDVINQISDSIRSRADEALL--VCKRIFG-IYRLRLRDNSEIHSLEE 816

QY 1182 -----VRDVSPEPVNAGMTVKDESLAPAVNPLVTPQKSGT-----LDNSLH-- 1223  
DB 817 AMERLYTLMSDDQKIVISYIDVNDKRLYEKIMEFPFDALGETQKXAKELKDAKCYIV 876  
QY 1224 KOLGAIISGLASPYTTSTPNINSVYVADSRGSLISTDSGNSLPERNSKNSLDRHQSS 1283  
DB 877 KOLIGEBNG-----NLREMKENGA--SNEAIATVEEMIEAITDETGRAQDM 922  
QY 1284 TLGNS-----VYRCDKLDQSEIKSLMCFYLILKMSGDDL---FTTWKASTSELDNF 1334  
DB 923 RASTSCRYGVGVQFRDRDHHNHL-----DEALEKFTYMLNEQKSQLK-- 968  
QY 1335 FTISVCLHQFOYMKRYIARTGMHARLQOLGSLDNLSTFNHSYGSADADVLHQSLLEA 1394  
DB 969 -TI-----YESDEBALHKKVMEF 986  
QY 1395 NIATEVCLTALDTLSLFTLAFKNOQLADHGNHPLMKVFDVYLCELOKHQSETALKVNF- 1453  
DB 987 FEAGAGLRASNASKKIYGVV--KRPRDHHNHLDEALEKYLTMINEBO-KSQMKTYE 1043  
QY 1454 TAYRSLIYKFEPTYE---GRADMCALCYELKCSNKLSTIRTEASQLYLFLRRNPD 1510  
DB 1044 SGDRBALYKQVLEFEAATGEVKEKAAV--ELKSACRHYIK-----D 1083  
QY 1511 YTG--KKSFVRTHLQVLIISQQLADVVIGIGTRFQOQSIIINNCANSDRILKHTSPSD 1568  
DB 1084 YIGDEKAEKIKEMESGVSTETI-----SKYDEFLAMTTDBKAKALRABASACK 1135  
QY 1569 VKDLTKIRITVLMATAQMKENHEDEMLVDLOYSLAKSVASTPELRKTMWDSMARHYK- 1627  
DB 1136 IYGVAKRRF-----RDHHEH-----NLBEALEKYL-----TWLNEQSQMKT 1174  
QY 1628 ---NGDLSBAMCYHNTALVAEYLTREKAVQMEPRPLPHSHSACLESRGV----- 1677  
DB 1175 IYESGDRBALYKQVLEFEAATGEVKEKAAVELK-----SACRHYIKDYIGDEKAEK 1226  
QY 1678 ---FROGCTAFRVIIPNIDSEAMMEDYQMVHFNEDVLMELBOCADGLMKABRYELI 1734  
DB 1227 IKEMESGVSTETISKYDEFLAMTTD-----EKAKALRABASACKKI 1270  
QY 1735 ADIYKLIPIYEKRDPERLALYDTLRAVSKYTEVNHSG-RLLGYTVFVAFGQAAQ 1793  
DB 1271 YGVAAKRL-----RDHHEHNLLEBAMGKYLSSWMSDEQAAQYKKIYGTGDRLATYNNKWE 1324  
QY 1794 YQFTSEFDVSGFEDEDECKEYTYKPKLTPISELISQLLKLYSKPNSGN--VKMID 1850  
DB 1325 -----LFESVPSDE--KEKATSQLKACRHYIK--DFIGKDNLAIVKEMER 1366  
QY 1851 SGKVPK-----DLDSKYAYIQ-----VTH 1870  
DB 1367 SGATNEAIGEKTIDEITAGLDQKQAOARASACKKIYGVKSRGRREHYEIDVDEAISK 1426  
QY 1871 VLPFDEKELDERK--TEFERSHNIRRFMEFPTQTSKQGVGEQCKRRTILTAIHCF 1928  
DB 1427 YLTWNEQKQKAIKQKXKDEKQOTIGCKKIMEFFELTSDDDEKARQKAA-----CX 1479  
QY 1929 PYVKGRIPVMT--OHTDNLNPIE--VAIDMSKQVAB 1961  
DB 1480 HYVK-----MYGGEKAAELKLLKQSGISLEMSKQKTE 1513  
RESULT 12  
YFP73\_CABEEL  
ID YP73\_CABEEL STANDARD; PRT; 4385 AA.  
AC Q09222;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Hypothetical 486.0 kDa protein B0228.3 in chromosome II.  
GN B0228.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhaditidae; Peloderinae; Caenorrhaditis.  
OX NCB1 TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Briseol N2;  
RA Leimbach D., Waterston R.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U23168; AAC3807.1; -  
DR WormRep; B0228.3; CB01744.  
KW Hypothetical protein.  
SQ SEQUENCE 4385 AA; 486003 MW; 50B8871B6C45FA23 CRC64;  
  
Query Match 1.7%; Score 184; DB 1; Length 4385;  
Best Local Similarity 17.2%; Pred. No. 0.086;  
Matches 436; Conservative 393; Mismatches 845; Indels 856; Gaps 118;  
  
QY 11 AETRKEFTRALSKPGTAELRQSVSEVRLAKPELLEPLDYENVYQK----- 61  
DB 174 AERVF-----AQLRASADEITIRVALGSMQLEQAFMLSLTSVRCGLRTI 223  
QY 62 -----KQIQLNDCLREMLFPYDFQRIARQGR--YICSTVPAKKEEQLFVTE 112  
DB 224 APSNIIANTEIFDYVAEEKMSV-----SGVMKRERKEHSSKKFTSSBEIIQGFWKGE 277  
QY 113 -----CIKTYNSDMHLVNYKE-----DYGSEF--RQLPNKV 143  
DB 278 RDEBRVVKLKD--RNVSPFHSLSVEAASSTSENVSLGMRKADQKSEFTISQKLSKEIV 335  
QY 144 -----KLDELPHYVYEVEDEVDKDEDAISGOKGITRGWL YKGMNMSAISVTM 194  
DB 336 SEAYGSESKLDQF--FQVMEKMDWSN--TELGEKE-----HTALSANI 375  
QY 195 RSKRKRFPHLIQGDGSKYKEEPLDKQEKSGIFGLYGVSPFNKKRRRFFELK--- 251  
DB 376 RT-----LAPROKVCDS--ILGKLKAPKODESVGTQIOETIRAT 413  
QY 252 --MODSSYLLADSEVEMEWITILNKILQLEFAAMQEKRGSHDEDEQSKLGSGS 309  
DB 414 AVMSVPAASLLITSSNSNFSKMSISEKAVFSNLIVT-----TSHNSTIGTSETTT 467  
QY 310 GLDSY--LPELAKSAREAEIKLSESRVYLFYLDPPAQKLDFFSASPEVKSFEKFKRI 367  
DB 468 ASISYODIPEMLASK-----LWISRSEKLEKEIRREPVIQTVESFW---- 509  
QY 368 LKCNCLSPLOCCVAVENERGPTTNVBPFFVTLSLDIKNRKISADFHYDLNHFVRQM 427  
DB 510 ---NTTNOEKVAVLNK-----IDSIYSSLNTLAASWENELISOELVRHTENAGKN 560  
QY 428 IATTS-ALMNGSGPETOSALRGILHEAMQYKOGIFSVTCPHPIPIVARIEKVLQGS 486  
DB 561 IKRISRELVSSTFKITSSDL-----QQPFNV-----LEKNDMSQ 595  
QY 487 ITHCAEPYKSSDSSKVAQVKNKAKQACQRLQGYMPFMAARTLFKQASGNLDKNAPF 546  
DB 596 I---SLPAQEHAIISKVRSLASASSELNIIQKLAPEMQEBETDLK-----IQAOQ 646  
QY 547 SAIRYDSKSLSDMDMLKLADFRKPEKAKLPIVILGNDITIDNVSDSPFNYSSTYP 606  
DB 647 ARVVLNVSSMSN-----TISSEETSRIP-----ENKATLTINY--GIIT 686  
QY 607 TKOFET--CSKPTFEVEEFVPCPKHOPYIITNHLVYVPKYLYKQSSQSAKARNI 664  
DB 687 TODLSLIGASSENQTIEMD-----YTEAKEELASGKLV--DKNLSVLKT--- 729

QY 665 AICIEFKSDSEEDSQPLKCIYGRPGGVFTSAPFAV-----LHHQNEPFEYDEIKIE 717  
DB 730 ---FIRESGDEVQGF-----NWTASDEKEVGAIIICEKUSIHHTLHTAIRTVIES 778  
QY 718 LPTQLHEKHHLTLTFPHYSCNDSKSGSTKKRPVVT-----QVGYSL 760  
DB 779 LSTDLQKEQKSLTHFFV-----KLST--RDVQAARISSEVDHLTLVEKQDWSKI 830  
QY 761 PLTKD--GRVTSQOHIPVSNLPSGYLVY-----QELGMGRHYGPEIKWVGSKPLK 812  
DB 831 DLVEPVYSSISNVQALATSSQSSGIGLKLPPAPEETSK---EIREINTAKCILN 886  
QY 813 ISTHLVSTVYTDQHL-----HNFQYQ 836  
DB 887 VISSINSTI--TSDASLDMISEQAI FSNIIIGWASNNLITSSQGTSSFGFNVFELSE 945  
QY 837 K-----TESAQAL-----GNELVKYLKSLHMGHVMIAF--LPTILNQLFVLVT-- 880  
DB 946 ARVLQESNKNQLIKVRESSEELIHGIMWSTASETEKVAIVKEKLETV--HQAMETLAIQ 1004  
QY 881 RATOEEVAVN-----VTRVLIHVVAQCHEGLESHLRSYKYAYKAE 922  
DB 1005 MATQ---SVNSLISSEGNLASIKSTLPTREVISAAGISNEVYQKULEVLTKEWMTI 1061  
QY 923 PVVASEYKTVHELTSTMTTLKPSADF-----LTSNKLKYSWFFPDVL 967  
DB 1062 SLPOSEY---ESISQVYALAEFPNCDSLIGKINAPQOTETTSHELIEKNTAAVAVN 1117  
QY 966 IKSMAGHL--ENSKVLLRNRPASVHHAVETVNNLMRP-----ITQKPDNPEA 1018  
DB 1118 VAAAEVSYSKSSSLAKL--PADEKALITNAGLVLSKDVSLCTSSSTFPDORI 1171  
QY 1019 SKNASHLAVFIKRCFTFMDRGFVFKQIINNYISCF---APGDKTLFEYKEFEELRVGNH 1075  
DB 1172 FPNQANISLGAVBETILQR--LREPIENQVQGWSTASNOEKSAFLAK----- 1219  
QY 1076 EHYIPLNIPMPFGKRIQRYQDLQDLSLTDFECRHHFLVGLLREVGCTALQEFREVRLI 1135  
DB 1220 ---QCLDTMYD-----TMKVI 1232  
QY 1136 AISVLKULLIKHSPDRYASRSHQARIATLYPLRGLLIENVQRINVRDVSFPFVNAQT 1195  
DB 1233 AIQ---MISQITIGDFSSASISSE-----TFKNDKA-ARBV--VAAEFG 1271  
QY 1196 VKDESL--ALPFAVNLVTPQKSTLDNSLHKDLGALISGASPYTSTPNISVNN--- 1249  
DB 1272 IANESVQALAEVLNVK--EWSDISLPEKLNQISSNVAVLD---SGVNCDSIIIGKUNA 1325  
QY 1250 -----ADSRGSLSTDSGNSLPERNSEKNSLUDKHQOQSTLGN 1288  
DB 1326 PEPQSAVYDQITSEHQTELVIANIKSAVSVISNDSVLKSE---DEKAVISNIANVL 1380  
QY 1289 VRCQDKLOQSEIKSLMCELYLTKMSDDAL-----FT-----YWNKAS 1327  
DB 1381 IISCDLASMSSTNSFEFORDLPEQONANVLGTPASQVPTRNLOEPIQSEVQGWSTNS 1440  
QY 1328 TSELMDFTIISEVCHQFOYMGKRYIARTGMM-----HARLQOLGSLDMS 1372  
DB 1441 SQE-----KASLVITQKLNKVKYDKAMKMAATEIAYQSINSINTSEENLESQSPDS 1492  
QY 1373 L--TNNSHYGSDDAVLHQSLEANIATEVCITALDTLSLFLAKNQQLADHGN--PL 1428  
DB 1493 IRESIMAFGVSESHV--QKTLIELISKSELGITVPAVEYEAATVONIKALSEPNCESI 1550  
QY 1423 MKC-----VPPVYLCPLOKQSEFRLAKNVFAL-----RSLTYK 1462  
DB 1551 LGLNAPPOQSAWMI---INMEQRKLTIVSISQISATISVINDSSLEKLTEDVDSVLK 1607  
QY 1463 PFTSYEGRADMCALC--YEILKCNKSLSS---IRTEASQLLYFLMR-----NNPD 1510  
DB 1608 ISEIIVSHDLTALSSCGDFNLQESNTQEKADIDLKTPNSGMLERLREPIEKOQGW 1667  
QY 1511 YTGKS-----FVRTHLOVITISVSQ-----LIADV-- 1536

```

Db      1668 STASNSEKQEMFLKEKVEITIHAMLTQFSASLSETVQRPDAVOSLALRSIOUTPREI 1727
Qy      1537 -----GIGTRFOQSLSIINN-----CANSDBLIGTGF-SSDV-----KD 1571
Db      1728 LCAAFGISSEQLEQTFRGJDEVDMSHIDIPCARLDLNLVNLAI VNNADANVFGNLISRPD 1787
Qy      1572 LTKRIRTVLMATAQMKHNDPEMLVDLOYSLAKSASTPELRKTLWDSMARIHVNGDL 1631
Db      1788 ENQGETSTVL-----EEKNRIGIPLNLQRLSDQTYQMTSSILTR-----SDEIIEVAVSNI 1836
Qy      1632 SEAMACYHVTALVAEYLTRKAVQMEPPLPHSHSACLRSGVFGROGCTAFVITYN 1691
Db      1837 -----ISLISSENL-----GDLISQ-----AVQLAQPN 1859
Qy      1692 IDEASPMEDVGMQDVHFNEDVLMELLECAGDGMKEREVELADYKLIPIYERGRPF 1751
Db      1860 MSDETEKSEFI-PRNLILNR-VVPESTSESIQSFWKTS--QLAEKASSTI-----S 1906
Qy      1752 ERLAHLYDLHRAVSQVTEVMHSGRRLLQTY----FRVAFQQAQVQFTDSETDV--EG 1805
Db      1907 EKLMLQ-----SEFVVSAAKQVSTSLTDVRRKIFPNQNSIIFGDLTRDVIYKA 1956
Qy      1806 FPEDEGKVIYKEPKLPLSEISQRLKLYSDKFG--SENVKMIQDS-----GKNPK 1857
Db      1957 FSVSEDTLTLFQSQLEQSDMSOI-----KLSSKQKSLISANIKSLATSNLDITLIGQLMR 2011
Qy      1858 DUDSKYAYIQVTHVIFPPEKELQERKT-EFERSNIRIRIMPEMPTQKQGGVEBOC 1916
Db      2012 EADSDQS-----EASLSEKSVEMENTFKIQIITD-----TVVKKFGSDEE- 2053
Qy      1917 KKRITLTAIHCEPPYK-----KRIPWVQHHTDPIEVALIDEMSKVAELRQC 1966
Db      2054 -RASLINLGLTSLVNLKAIQVASEERIPDLSIDSTASTTPVKIOENHROQTEGVQGF 2112
Qy      1967 SSAB-----VDMIKQLKQGSVSQVAVAGPLAVARFL-----DDNTKRYPD 2010
Db      2113 WSKHEHPNETVAVNRKIEVLKSI--LNCYAVAECHKSLERSLEKPSQODCATRGVT 2169
Qy      2011 NVV-----KLKEVFR-----QVEACGQALAVMERLI-----KED 2041
Db      2170 EYVQNAFAVTEQSYKILKVLDSMONFEIPEAVKETISKLKILNLPVDTLLVSKYKS 2229
Qy      2042 Q-----LEYOE--EMKANYR-----EMAKELSEIMHEQICPLEEKTSLVPLSHIF-- 2085
Db      2230 HYSKVEBELVQFMFNFKSALNSEIMQNSK-----C-IEETVVLPLVAILASA 2282
Qy      2086 NAISGTPST 2095
Db      2283 NLATIVPTES 2292

```

## RESULT 13

```

A11_HUMAN
ID AK1_HUMAN STANDARD; PRT; 1901 AA.
AC 09UKA4; 075124; GN0K7;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE A-kinase anchor protein 11 (protein kinase A anchoring protein 11)
DE (PRAK11) (A kinase anchor protein 220 kDa) (AKAP 220) (hAKAP220).
GN AKAP11 OR AKAP220 OR KIA00629.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20323159; PubMed=10864471;
RA Reinton N., Collas P., Haugen T.B., Skahnegg B.S., Hansson V.,
RA Jahnson T., Tsekken K.,
RT "Localization of a novel human A-kinase-anchoring protein, hAKAP220,
RT during spermatogenesis."

```

```

RL Dev. Biol. 223:194-204 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1290-1901 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9840380; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176 (1998).
RN [4]
RP SEQUENCE OF 1121-1469 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE A
RC -I- ANCHORS/TARGETS THEM.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC IN PREMEIOTIC PACHYTENE
CC SPERMATOCYTES AND IN THE CENTROSOME OF DEVELOPING POSTMEIOTIC GERM
CC CELLS, WHILE A MIDPIECE/CENTROSOME LOCALIZATION WAS FOUND IN
CC ELONGATING SPERMATOCYTES AND MATURE SPERM.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LUNG, LIVER,
CC KIDNEY, TESTIS AND OVARY. WEAKLY EXPRESSED IN SKELETAL MUSCLE,
CC PANCREAS AND SPLEEN.
CC -I- DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
CC HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF176555; AAF07045.1; -
CC EMBL; AL136527; CAB89419.1; -
CC EMBL; AB014529; BAA31604.1; -
CC EMBL; AK002166; BAA92117.1; ALT_INIT.
CC DR Genew; HGNC:369; AKAP11.
CC MIM; 604696; -
CC FT DOMAIN 1650 1663 PKA-RII SUBUNIT BINDING DOMAIN.
CC FT DOMAIN 1705 1767 SER-RICH.
CC FT CONFLICT 1303 1303 K->R (IN REF. 4).
CC FT CONFLICT 1444 1469 LDPYRNEVSQLYSFTSLVSHITKDA -> CGPSVELSPWK
CC FT MOTRCGRGNSWKR (IN REF. 4).
CC FT MOTRCGRGNSWKR (IN REF. 4).
CC SQ SEQUENCE 1901 AA; 210510 MW; 089331C46A75672 CRC64;

```

Query Match 1.7%; Score 183.5; DB 1; Length 1901;

Best Local Similarity 18.7%; Pred. No. 0.026; Indels 581; Gaps 84;

Matches 337; Conservative 246; Mismatches 637; Indels 581; Gaps 84;

```

Qy 98 PAKAEEAQSUF-----VTECI-----KTYNSDMHLVNYKYEDYSGEF 135
Db 436 PKASAREDSGLFSPIRSSAFSPGLGCTPAFCFCQDIDGDAIHENHDSVYTYEDYA---- 492
Qy 136 RQLPNKVVKLDLPLVHYVEVDEVDKDDAASLSQKQIGITVGGMLYKGNNSAISVTWR 195
Db 493 -----KSISCEVLGSLVLRTHHTNTLSNINSIKHGKNTVTFPGHNI.DQKNKSQKSLMIK 547

```

QY	126	SPKRRF-PHLIOLGSGSYKFEFLKLOKEPKSITLGLYGSFRANNVRRRAFLKQD	254
Db	548	DSIQFAADLVKESFGS---AFKLOK-----GVSSCTNL-----	580
QY	255	KSSYLLAADSEVMEEWITILNKILQLFBAAMQKRNDSHEDDEOSKLEGGSGLDSTY	314
Db	581	---YHLA-----IKLTSVLOMAFDELRRQ-----AFGLKEALISGLANF	618
QY	315	L--PELASARAEIKLKS--ESRYKFLPLDDAOKLDF-----SAAPEVK	357
Db	619	LVSEALSNALKOLQYVKQIFNTVARPAADL-ABELVFEGIMEVCOFSYPOTPASPOCG	677
QY	358	SPEEKGRILVKNDLSFNLOCCVAENEEGPTTWEEPFVTLSPFDIKYNRKISADPHV	417
Db	678	SPDFE-DKVVUKYADLSESV-----IQCAFILSLOVDVTFTTKAAVSST	722
QY	418	D-LNHFVSROMIATT---SPALNMGSPETOSALRGILHEAAMQYR-KOGIFSVTCPH	470
Db	723	DNIKVYASBVSPYTOAVTFSPSPFNQOIMYTKP---VGEYKKEYTVQOALFCTS---	774
QY	471	PDIPLVARIKLOOSI-----THC-ABEYMKSSDSK-----VAQKULKAKQ	513
Db	775	---GIVTISVPPLASALLPYHISTTACQAKHLSSDSNSNGSAQYHIATK-NREKA	830
QY	514	ACOR---LGQYMPAPMAARTLFQDASGNLDGNARFSAIYRDSNKLNSNDMLKLLADR	570
Db	831	ACLRIKICPSEHN-----GN-----QNDKPFINDI-----EMQ	860
QY	571	KPEKAKPLVIGLNDITIDNVSSPPYNAVSSYPTQOFETCSKPTPTFEYEEVPCIP	630
Db	861	SSSKLPNDPAIISNSAAVHT-----IVNETLBSMTLSVTKWDEKTDVLT	908
QY	631	KHTQPYTYTNHLYVYPKLYKYDSQSKFAKANIMACIEFKDSDEDSQPKICYGRPGG	690
Db	909	KSLKKTPEFSH-----C-----DAVLQCSASSNK	935
QY	691	PVFTRSAPAAVLHHQNPBYDEIKIELPTQHEKHLLLTFHVSCNSSKSGTKKRDV	750
Db	936	DMFAURLSKSIHKHS---IDKSKVIP-----NIDKNAAVKES-	970
QY	751	VETQYGSWMLPLKXGRVATSEQ--HIVSANLPSGYLQGYOELMGHRYGEIK--AWDG	806
Db	971	-----LPVSGESOLTPKSKPKFPDSQ---QLTHCSLSAAKOCBECKVSMVHG	1017
QY	807	G-----KPLK--ISTHLVSTVYTQD---HLHNFQYQ	836
Db	1018	SSLFTLPCSPAVTQKSDLKESAKQOPLKKNLNTSLALSFGDENPFPHSTFSSYAL	1077
QY	837	KTESGAQALGNELVYKLSLHAMEGHVMIATPLTLNQLFRVLTATQBEVAVANTRII	896
Db	1078	TCVDDLHVEDKQYR-----DENV---IPDTPSTPLVPSASSEMDIKLTKLXK	1125
QY	897	HVVAQ-----CHEEGLESHRSYVYKAPYAVASEYKVVHELTGKSMTILKP	946
Db	1126	GELAKEPARATPSTPHNHSVGS-LSBNBQNTIEKEBFMLKLMRLSEVESSSEGEL-P	1183
QY	947	SADFLT--SNKLKY--SWFFDVILIKSMQHL-----IENSKYK--LLRNQ	988
Db	1184	EVDVSESHSGKKVCQAEALATILISLATEMASHDNKIIOBPKNCLINQVOSQRSVP	1243
QY	989	-FPASVHNAVEVVV---MLMPHTQKFRONPREKSNANNSLAVFIRCTPFMDRGVVF	1043
Db	1244	TFPLNSDENLTKLNCNPAQDLAAEVIT-----EAEKIAK-----VENC--MLF	1283
QY	1044	KQINNYISCFAPGDEKTLFEYKFEFLRVYCNHENYI-PLNLPMPG--KGRIGYODLO	1099
Db	1284	KOKRY--SCYADGDEDDYKEEKLDIEAVV--HPRVDPFIIISLPSSCGSMGMYKPECE	1339
QY	1100	LDYSLTDFEFCRNHPLVGLLREVGVALDQFREVRGLAISVLKNLILKXSPDRYASRSHQ	1159
Db	1340	---SVTDEYAGH--LIQILKQEGNS-----ELM-----DOYANR--	1370

```

QY 1160 ARIATLVLPLFGLLIENVRINRA-DVSPFVNAGMVTIKESLTL-----1203
Db 1371 ---LAYSVSGLQEAARKTKTKVCNSRMFVPSQVTKTNELLMFSKHNHOADKKQ 1426
QY 1204 -----PAVNDLVTPQKSGST-LDNSLHKD-----DLGASIGIASPYTT 1239
Db 1427 SKRNBGVCFCKNOTCBERTLDPYRNEVSQLYSSTSLVHSITLDAKEELTASLVGLPKSLTD 1486
QY 1240 S-----TPNINSVRNADSRGSLISTDSGNSLPERNSEKXNS 1275
Db 1487 SCLFEKSGSEEDNECHVTPELPKSLQSSQNHRRFVHTSGSLNGYCGCDNVQAVEGYAKK 1546
QY 1276 LDKHQQSSTLIGNSVYRCDXLDQDSFKSLMCFYLIILKSMDDALFTYMNKAS--TSELM 1333
Db 1547 VDDTLLELTIGSTVFRVSETTKSGADR-----VTYAEKLSPLTGQACR 1588
QY 1334 FFTISEVCLHOFQYMGKRYIARTGMM-----HARLOQSLDNSLTFPNHSYCHS 1382
Db 1589 YCDLKE-LHNCTGNSSSHFFRQSLASSKRASPKESSRYQK-----SRIFHLSPQI 1640
QY 1383 DADVLHOSLEBANIAATEVCLTALDLSLFTLAFKNOILLADGHNPMLKKVFDVYLCPLOK 1442
Db 1641 HVNLDKRAVLAEKTVAEABIEKARELSTSLA-----ADSG-----IQEGASFAES 1687
QY 1443 HQSET---ALKQVFTALRSL--LYKPESTFPEGADMCALCYTL-KCCNSKSSIRTE 1486
Db 1688 LATETMTAAVTNVHAAVSSKKEIDFQSTESVSQQWNLISGDSSTGSWSNLSFEDEHOD 1747
QY 1497 ASQLLYFLMRNPFYTGKKSFVRHQLQVILISQGLADVVGIGGTRFQOSSLINNCAN- 1555
Db 1748 ESSGFHLSEBNGSSMS-----LGLBEDLYEDNISFTSDSG 1788
QY 1556 -SDRLIKHTSPSSDVKDLTKRIRITVLMTAQMKENDPEMLVDLOYSLAKSVASTPEL- 1613
Db 1789 PDDKDEEH--EDVEGELGQGGKTLTLTNIDMEPCVTDPQRIILQWLIA-SEAEVAVELY 1844
QY 1614 -----RKYLDMSMAKIHVNGLSSEAPACYVATLVAEYLTRKEAVQ 1656
Db 1845 FHDSANKEFMLSLOLEKGV-----KVGDLQAVLQYEVMEKASSEERCKSLFD 1895
QY 1657 W 1657
Db 1896 W 1896

RESULT 14
ACF7 MOUSE STANDARD; PRT: 5327 AA.
AC AC ACF7_MOUSE P97394; P97395; P97396.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (microtubule actin crosslinking factor) (MACP).
GN ACF7 OR ACP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=BALB/c;
RX MEDLINE=20069791; PubMed=10601340;
RA Leung C.L., Sun D., Zheng M., Knowles D.R., Ilem R.K.H.;
RT "Microtubule actin cross-linking factor (MACP): a hybrid of dyctonin and dyctophilin that can interact with the actin and microtubule cytoskeletons.";
RL J. Cell Biol. 147:1275-1286(1999).
RN [2]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97124842; PubMed=8954775;
RA Bernier G., Mathieu M., De Repentigny Y., Vidal S.M., Kothary R.;

```





QY	330	SSABEYVUSPEBKPEKRIILVKNCDNSFNLQCCVNAENEGPTTNBPFVTILSPITKXNR	403
Db	1594	KDLOQDOSHSTSPATAV-----DIEGLEBNO-----T	1623
QY	410	KISADPHDNLHFYSYROMIATTSPALMNGSGPETSALRGILHEAMQYPKOJIFSUTCP	469
Db	1624	KLS-----POELTALREKLNQKEYE-----	1645
QY	470	HPDITLVAR---IEVLOGSITHCPEPIKMSDSKVAQKYLKAKQACORLG-----	519
Db	1646	---VLOERTVQAQLEBAVTSALQ---QETEKSKATTELAENRKIDALLDWYLLMG	1698
QY	520	---QYRMP-AMA-----RTLFKDASGLDGN-----ARFSAIYRD	553
Db	1699	HLRSHRLAFOQMSSLEAMEKOTLATDGHVDVNOUPETILDRQYELMKARHGEILSQ	1758
QY	554	SNKL---SNDMLKTLADFKKPEKMAKPYLGNLDTITDNVSSDPFNVNSYIPTKQ	609
Db	1759	QNFIVATOSAFSFLDOHSHNLPEEROKLOEKGELK-----EQYAA5-LAR5E	1806
QY	610	FETGSKDPTTFEEVEFPVPCIRPHQNPYITNHLVYVPKYLUYS-QKSPKARNIACI	668
Db	1807	AEKKTQALRDELQFL-----ODKHEFENWLOOSENEL-DSMHKGS5PEALNSL	1855
QY	669	EFKDSDEDSOPKLKCIYRPGGPFVTRSAFAVLHNNHNP-----FYDEIK-	715
Db	1858	KRQGSFSED-----VISHKGDLRFVTISGQVLFTENNFBQGPATRNILVMEKDA	1911
QY	716	IELPTQLEK-----HHLTLT-----FHVSGDN-----SKSGTKR-RDVETQNG	756
Db	1912	TERYTLTSHSKIRGLSHLSMLIGYOQFOSSADSIQAWLTCEBAVGKLSLDTVA5DGV	1971
QY	757	-YSWPLRLKDKGRVYVSEOHIPVSAVLPGYLGYOELGMRHYPEIKWVDCG-----KPL	810
Db	1972	LQOQATTKQJQOELAEHQVPY-----EKLOQKAH-----DLIDIEEPALDCRPI	2017
QY	811	LKISTHVLVSTVYTODQHLHNFQOYCOQKTESGAQALGNELVYKLSLHMEGHMIAFLPT	870
Db	2018	QETDTSISSRFONLSCSIDERSALLOKAIQASQSOVSEMSILSOSIREVEONL-----	2070
QY	871	ILNQLFRVLTRATQOEEVNVNTRVLIHVVAQOHEGSEGLSHRSYKYVYVKAPEVYASBYK	930
Db	2071	-----ERQOVASLSSGV-----QALANNMKLKODIARO-----KSLE	2105
QY	931	TVHEELYSMTILKPSADFLTNNLTKLYSWF-----FDVL	967
Db	2106	ATHDWATRFMETADNSASAVL-QGKLAEISGFQOLQOQOBKESNLKCLLQAEMLFQOL	2164
QY	968	IKSMQHLIENSKVULRNQRPBAYVNHAVEVYVNMMLPHITQKRDMPKASKANSHLA	1027
Db	2165	SNKLQOEFENKSRKLASGNQPDQDIHFSQO-----IQEITLAMEDKENDLTLER-LV	2217
QY	1028	VFIKRCPTFMDRGFVFKOINNYISCEFA-----PGDPKTLPEYKEEFLVNCNHEHY	1078
Db	2218	TLIGSCGFALDLSQHQDKIQNLKKQFTELQTVQERKEDASQCOQOBLDFKRLIRTPQK	2277
QY	1079	IPL---NLP--MPPEKGR-----IQRYODLOLDYSLTDFCRNHPVLGILLIREV---GTA	1125
Db	2278	LKETEGNVPKATFVSAKELEKQIEHLKDLISDWSK-----GALLGEIYAKGTA	2327
QY	1126	LOEFREVLIAISVKNLKLKHSFQDRVYASHHORIATVLPLYGILLIENVOYRNV-RD	1184
Db	2328	LE-----SLMDITAPDS-----QAKTGST-LPPVGS5VGSVNGYHTTKD	2366
QY	1185	VSP-----FPVNAQ-----MTVQESLALPAVNPILVTPQKSTLIDNSLH--KDLIGAI	1230
Db	2367	LTEIQCMDFDVSKEKLEMLVLRHQESLQYFVSMEVQKXASVQLQWLSKEBEVLKAM	2426
QY	1231	SGIASPYTST-----PRINSVRADSRGSLISTDSGSLPERK5E	1271
Db	2427	DATUSTYETETEVKQAQESNKAFLAELQONSPIQCVKCA--LAGLTKYPN5QAEANWK	2484

QY	1272	KSNSLDHKQSGSTGNSVVRCDKIDJDOSEIKSLYKSCFYLIKSMGSDALFTYWNVASTSEL	1331
Db	2485	MOEDLNSMEKAT--EVTYAROKOLEES--ASHLACF-----QAASQL	2524
QY	1332	MDFFTISEVCL-----HOFYWKRYIARTGMMHARLOOL--GSL--DNSL	1373
Db	2525	RPMIMEKEIMMGVLGRPLSIDENMLKQOVQFWLKEFEFARR--QOHQLNEAAGILTCGDM	2583
QY	1374	TFNHSYGHSDADVLHOSLLEANIITEVCLTALDPLSLFTLAFKNQOLADHGNPLMKVF	1433
Db	2584	SPSASQVHKDQGISQKQV-----LT--DKLNSRSQIDQATYSTQYODLQDS	2633
QY	1434	DVYLCLFQKQSGETALKNVFTALRSLLYKPESTPEGRADWCAALCYEILKCSNKLSSI	1493
Db	2634	EKVATIGQRLSGGSAISTQPEAVAKQOL-----EETSEI	2666
QY	1494	RTEASQLLYPMRNPNFYTKGKSIVVRHLOVITISVSQLADVIGIGTRFGQSSILNNC	1553
Db	2667	RSDIGQL-----DNEIKEA-----QTL-----C	2684
QY	1554	ANSRLIKHTSPSSDVKDLTKRIKTYMATQMKHENDEPMVLQYSLAKSYASTPEL	1613
Db	2685	QELSLILEOYLKQELK--KRETLTVLPLQGL-----DLADNRSLQALALSTQOF	2735
QY	1614	R-----KTWIDSMARJHVKNGDLS--EAMCYVHTALVAEYLTEREAVQMERPLLPH	1664
Db	2736	QOMFDELRTWIDDEKQSOQAQKCPISAKLERLQCL-----QENEEFQKJNINH	2783
QY	1665	SHSACLRSGRGVFRGCGTAFRVITPNIDEBASMMEDVGMDVFNEDVLMELL--EQC	1721
Db	2784	SGSEIVLVAEG-----EALLSVPRQEEKTKTQNLQVLKRSW--EDLSKTKANRSGRL	2835
QY	1722	ADGMLKAEVLYADIYKLIPIYE--KRDPERYL-----AHLYDTLHRAVSKTEVNHSG	1775
Db	2836	KDCMQKQKQYQ--GHVEDLVPMIDECKSKMPELQVTLDPVQLSESLRSKAMLENA--EK	2891
QY	1776	RRLIGTFRVAFFQQAQYQPTBETVEGFEEDG----KEIYYE--PKLTPLSEIS	1829
Db	2892	RRSL-----LEINLSADILINSEIDEDEIRDKAGINQMDAITBELQAKTSSLEMT	2946
QY	1830	QRLKLKYSDKGSENVKMIQDSGVNPKJDLDSKAYUIGVTHVLPFP-----EKELQ	1881
Db	2947	QRL-----KEPQESFNIEKKYVCG-----AHQOLEIFDALSQASNNKMLE	2987
QY	1882	ERKTEPE-----RSNINIRPEFMPEPPTQTKGQSGVEBOCKARTILLAINCPYVKKR	1934
Db	2988	KLKAKQOEVLQALBPQVUULRF-----TQGLVEBA-----PDGSDA	3033
QY	1935	IPWVQHNHTDLPNTEVALIDEMSKKVAELRQSCSAEVDMLKLOLQOSVAVQVNAGPLA	1994
Db	3024	SPVNHQ-----AEVAQDEF-----LEVKQAVSS--CLTMENKLBGIGQFNKRVEMF	3069
QY	1995	YARAFLLDT-----NTKRYPDNKVYKLIKEVFRQFVEACGQALAVNEBLIKEDQLE---	2045
Db	3070	SQLADRLDELDGMAIGRDTBDSIQSGIEDV--RLFLNKI--QALRPD---IEDSEAECKRML	3124
QY	2046	QEEKAKYUREAKELSEIMHEQICPLEKTSVLPNSLHI	2084
Db	3125	EEEGTLLDLGLKREL--EALNNQSGCKLTERGKVRQOEL	3162
RESULT 15			
YD86_SCHPO			
ID	YD86_SCHPO	STANDARD;	PRT; 1957 AA.
AC Q10411;			
DT	01-OCT-1996	(Rel. 34, Created)	
DT	01-OCT-1996	(Rel. 34, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Hypothetical protein ClF3.06c in chromosome I.		
GN	SPAC193.06C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		



CC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RK MEDLINE=21048401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Pat N., Hayles S., Bigham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Frazer A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsis K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ocell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynopreuz B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fricz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Zimmermann W., Medler H., Wandut R., Furtelle B.,  
 RA Egger P., Zimmermann W., Medler H., Wandut R., Furtelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revelta J.L., Moreno S., Armstrong J., Forzburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovskii G.V., Useery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880 (2002).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; Z70690; CAA04624.1; -  
 CC DR Hypothetical protein.  
 KW SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;  
 SQ  
 Query Match 1.7%; Score 181.5; DB 1; Length 1957;  
 Best Local Similarity 17.4%; Pred. No. 0.035;  
 Matches 369; Conservative 326; Mismatches 803; Indels 623; Gaps 83;  
 7 LPSAETRTKTRALSKGTAELRGSSEVVRGSVLLAKPKL-IEPLDYENV----- 57  
 176 LTSEEDVSYFOKKLTMTESNFAKQSEADLSROLTYTHKDKKEDYKEDVSSIK 235  
 58 IVOKTQIINDCLR-----EMLLFPYDFOGAILRQG--RYICSTVPAKAE---BA 105  
 236 ASLAEQANSKSLRGORLEKLVSSNKTIVSTLRQTENSLRACEKTLQDEKKAINEB 295  
 106 QSLPTETCKIKTNSDMHLVNYK-----YEDYGEFRQLPKNKYKDKLPVAYEVD 156  
 236 DSKLLEEL-----KINVANYSDAIVHKDKLIEDLSTRISFDNMLKSEEDTLISKNKE 349  
 157 BEVDKEDDASLGOKGITTKGWLKGMNNSAISTVRSFK--RRFPLIOLGDSGYF 214  
 350 KLLRN-----TIGSLKDSRTS-----NSOLEBEMVYLKESNRTIH-SQITDSESL 394  
 215 EPLKDLQKSPKGSIFLGLYGVSRNNKYRPAFAELKMDKSSYLLAADSEVEMEWITI 274  
 395 SSFPEENKSLKGSID-EYQNNLSKDKMYKOVSSQLE-EARSLAHATGKLAINEERDF 452  
 275 LKTIQLNEEAMQER---NGDSHEDDQSKL-----EGSGS 309  
 453 QNKKIK-DEKIBODLRACLSSSNELKERSALIDKQDELNNLRQIKQKKVSESTOS 511

QY 310 GLDSYLPFLANLSAREAEI-----KLKSESRVKLFYLPDPDAQKLDPSABPEV----- 356  
 DB 512 SIQSLQDILNEKKHGEVYESQLNELKGELQTEISNSBHSLSQSLTAAEKAAVAATNNE 571  
 QY 357 -----KSEERFGRIILVKCNDLSFNLOCCVAENEBGTTTVVPPFVLSLFD 404  
 DB 572 LSEKNSLQTLGNAPQOEKLAWSWQ-----LKENEQ-----NFSILD 608  
 QY 405 IKYNEKISADPHVDNHSVRQMTATTSPALMNGSGPETSALRGILHBAAMQPKGIF 464  
 DB 609 TSFKLNSHQELENHQTITQKLDTS-----SKLQQLQERANPEQKESTL 656  
 QY 465 S-----VTCEHPDIFVARIKYV-LOGSITHCAEPYKSSDSKVAQYKXAK 512  
 DB 657 SDENNLDRTKLLKEBSKXSLIKQGEDVDSLEKNIQTLKEDLRKSEBALRFSKLAQILR 716  
 QY 513 QACQRL-CQYRPPFMAARTLFKDASGULDKNRARSATYRQDSNLSDDMLKLADF-- 569  
 DB 717 EVIDMLKQKHE-----TLEQRNDLHSLSDAKNTNALSLELT-SEBDVRLRANVET 770  
 QY 570 --RKPEKAKLPIVIGNDITIDVNSDF-PRVNVSSYIPTRKQFETCSKTPTPRABEV 626  
 DB 771 LTQSKANKQSTLSVNSYQSIINLYHELRDDHV-----MOSQNTLIESSKLKT 822  
 QY 627 PCIPHTQPYTYITNHLVYPRYLKYDQSKSPAKARNIATCIBFQDSDEBDSQPLKTYG 686  
 DB 823 DCENLTQGNMTLIDNVQGLMKHKNVQSKSVLSKEVNGKLSIDLKN----- 868  
 QY 687 RFGGVPRPSAPAAVLHHQNPEDYDIKIELPTQJLHEKHMLLTPFHVSCNNSKSGTK 746  
 DB 869 -----RRSLNVAISDND-----QILTQJLAE-----LSK 892  
 QY 747 KRDVETQGVSWPLLDLNGRVVTSQHIPIVSANLPSGYLGVELGMGRHGEPEIKMVD 806  
 DB 893 NYDSLEQ-----SAQLNSG-----LKSLEA 913  
 QY 807 GKPLKISTHVLSTVYTDQHLH---NFOYCKTESGAQALGNEL-----VYLYKSL 856  
 DB 914 EKQLL-----HTENBELHRIIDLTLGTGLKIEBSKSDGLKLTARQEIENLKEB 963  
 QY 857 HAMEGHWIAPLPTLNLDFVLTFRATQEBVAVNVTYIIVHVAQCHESGLSHLSRYVK 916  
 DB 964 NMSQQA-----ITSYKSLDETLSKSKSLKEDIE-----HLKNVSEVEVERNA----- 1008  
 QY 917 VAYKAPVYASEYKTVHBEKLTSMWTLIKPSA--DELTNKLKXSWPPFDLI----- 968  
 DB 1009 -LLASNERLMDLKNGENINIASLOTETIKGRENDLOS-KLSVSVSEYENLLISSQTN 1066  
 QY 969 -----KSAOHLIENSRYKL--RNQRPASVYHNAVETVVMMLMPHIYQKPRDNEASKN 1021  
 DB 1067 KSLBEKTNQKTIKNGVQKLDKQQR-----NVELEBELTSKYQKQGEENAQ 1113  
 QY 1022 ANHSLAVFIK-----CFTPMDRGVFVKQIINNTISCPAPADPRTLBYKPEPLRVN 1074  
 DB 1114 IKDELLARKKSKQOHDCANFVDD-----LKEKSDALEOLTN 1151  
 QY 1075 HEHYIPLMLPMFPGGR--IQRYQD-----QLDVSLTD-----EFCRNPIVGL 1117  
 DB 1152 EKNEELIVLEQSNNSNEALVEERSDLNRLSDMKSSLSDSQNVISVIRSDLVAVDELD 1211  
 QY 1118 LIREVGTALQEFREYRLAIVLAKUL--IKSFDDRVAISRSHQARIVTLPLPGLLLEN 1176  
 DB 1212 LKQKQDSISTQSVSCQORDLDBLKKCEBSFNKYNVANSBELCTKSHIDVVSLELDN 1271  
 QY 1177 VQRIINRVDSPPVNAQ-----MTVQESIALPVPVLTPOKGSTLDNSL----- 1222  
 DB 1272 -----FVFAGNFSELSRLTVLSLNVLDADFQV--NFKQMEIDNLTJTTDAEF 1318  
 QY 1223 -----HNDLGAISGLASPYTTS-----TPNHSVRNADR----- 1253  
 DB 1319 TKVVADLEKLQHEHDBWLIQRGDLEKALKDSBKQFLRKAEAMTENIHSLEBGETTKKEI 1378  
 QY 1254 GSLISTDSGNSLPERNSEKNSLDRHQOOSTLGNSVVRCDK-----LDOSEIRKL 1303

```

Db 1379 AELSSRLBNOL--ATNKLNOQLDHLNOEIRLKEDVLKESLIISELSISNOQKES8 1436
QY 1304 LMCFLYILKSMDDALFTYNNKASTSELMEFTISEVCLHQFOYMGRIYARFGMMHARL 1363
Db 1437 LIDAKNELEHMLDD-----TSRKNSSLME-----KIESINSSLDOKSFEIASAV 1480
QY 1364 OOLGSLDNLSTFNHSGHSDADVLHOSLEANIATEVCLTALDPLSLFTLAFRN----- 1417
Db 1481 ERLGALOKL---HSELSISLMEINIKSLOLQAKKEKIQVESTIOELDHETIASKNNEGKL 1536
QY 1418 ----QLADHGHNPILMKYVDVYLCLQKHQSETALKNVFTALRSLIYKFPSTFYEGRAD 1473
Db 1537 NDRDSIIRDLSEN--IEQLNNLLA-----EERSAVYRLSTERESEILOP----- 1578
QY 1474 MCAALCEILKCNKSLSIIRTEASOLLYFMNNNDYTGKSFVPTHLOVIIISVQLTA 1533
Db 1579 -----NSRLADLEYHKSQV-----ESELGRSKLKLASTTEEL-- 1610
QY 1534 DVVIGIGTRFQOSLSIINNCAUSDRLIKHSPSSDYVDLTKRIRTVLMAATQKHEHNDP 1593
Db 1611 -----QLAENERL-----SLTRMLDLQNOVKDLSNIKOSISEDLRTL 1648
QY 1594 EMLVDIQLSLAKSYASTPELRTKWLDSMARIHVKNGDLEAMCYVHTALVAEYLTRKE 1653
Db 1649 RSLIEDSVASLOKCKIKSNTVESLODVLTSVQARNALBD-----EVSRSYDKIRRRDD 1702
QY 1654 AVQWEPPLPHSHSACLRSGRGVFR---QGCTAFRVTIPNIDEASMMEDVGMODVHF- 1709
Db 1703 RCEHLSGKLKULHSQ--LEEQHETFFRAEQORMTQLOGLKETVKKQKELKULNROBOLI 1761
QY 1710 -NEDVIMELLEQCADGLMKAERYE-LIADIYKLIPIYERKRDPERLAHLVDTLHAYSK 1767
Db 1762 PRESSIIV-----YESYIRDEKEIIVLOERLNGIE---LSQOLPKGY-- 1800
QY 1768 VTEVMSGRLLCTYFRV-----AFPGQAQYQFTDSETDVGEFPEDEBGEYIYK 1818
Db 1801 -----FGYFFKTRNREMEVLDSFKQYAKLOFLAGAEFIYKFKEDLEKCAAEK 1849
QY 1819 EPKULT-----PLSEISORLLKLY-----SDKFGSENVKMIODSGKYNPKD 1858
Db 1850 EKQATFDNTSEKVENIGKSIEALYFALNREISFRKSLALSKSAYHNL-LVRDSPKFNP-- 1906
QY 1859 LDSKVAYIQVTHVIPPEDEKE 1879
Db 1907 -DS-----QITYSIPVTNTKQ 1921

```

Search completed: July 14, 2003, 18:16:35  
 Job time : 46.9153 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:07:56 ; Search time 97.7492 seconds

(without alignments)  
4441.379 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936  
Sequence: 1 MSQPLPLPASAEIRKFTPL.....ISGTPSTWVGMTSSSSSV 2107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8002	73.2	1539	4	Q9B229	Q9B229 homo sapien
2	4605.5	42.1	909	4	Q9B227	Q9B227 homo sapien
3	4491	41.1	878	4	Q9B228	Q9B228 homo sapien
4	3693	33.8	724	4	Q9B226	Q9B226 homo sapien
5	3305.5	30.2	1984	5	Q9VM04	Q9VM04 drosophila
6	2888	26.4	738	11	Q63603	Q63603 rattus norv
7	2739.5	25.1	881	4	Q96MN3	Q96MN3 homo sapien
8	2192	20.0	448	4	Q9B225	Q9B225 homo sapien
9	2180	19.9	611	4	Q8WX10	Q8WX10 homo sapien
10	2094	19.1	417	11	Q921Y6	Q921Y6 mus musculu
11	2085.5	18.4	425	6	Q9GM45	Q9GM45 macaca fasc
12	2007	18.2	696	4	Q9BWX9	Q9BWX9 homo sapien
13	1985.5	18.1	534	4	Q8WX11	Q8WX11 homo sapien
14	1977.5	17.3	2374	5	O45377	O45377 caenorhabd
15	1867	17.3	1538	4	Q9H1Q3	Q9H1Q3 homo sapien
16	1887	17.3	1540	4	Q9H1Q2	Q9H1Q2 homo sapien

17	1785	16.3	542	4	Q96BY6	Q96BY6 homo sapien
18	1759.5	16.1	2284	5	Q8TIC8	Q8TIC8 dictyosteli
19	1733	15.8	500	4	Q9NIX8	Q9NIX8 homo sapien
20	1680.5	15.4	1782	5	Q9VPI9	Q9VPI9 drosophila
21	1664.5	15.2	2018	5	Q20487	Q20487 caenorhabd
22	1656	15.1	1373	4	Q8TEP1	Q8TEP1 homo sapien
23	1625.5	14.9	1628	4	Q9P2F2	Q9P2F2 homo sapien
24	1593.5	14.6	1302	4	Q9C092	Q9C092 homo sapien
25	1523	13.9	1180	4	Q9H7P2	Q9H7P2 homo sapien
26	1492.5	13.6	2621	5	Q8SSW5	Q8SSW5 dictyosteli
27	1395	12.8	415	4	Q9VM06	Q9VM06 homo sapien
28	1230.5	11.3	860	4	Q96HP0	Q96HP0 homo sapien
29	1226	11.2	595	4	Q75178	Q75178 homo sapien
30	1188	10.9	844	11	Q9CUB3	Q9CUB3 mus musculu
31	1141	10.4	624	4	Q8TB82	Q8TB82 homo sapien
32	1128	10.3	626	4	Q96NG5	Q96NG5 homo sapien
33	1084	9.9	444	6	Q9BE49	Q9BE49 macaca fasc
34	1043	9.5	1830	10	Q8SAB7	Q8SAB7 arabidopsis
35	1034	9.5	567	11	Q8RLA4	Q8RLA4 mus musculu
36	975	8.9	569	11	Q8VDR9	Q8VDR9 mus musculu
37	945	8.6	513	4	Q8WUY2	Q8WUY2 homo sapien
38	941	8.6	741	4	Q96N67	Q96N67 homo sapien
39	935	8.5	513	11	Q9DBQ2	Q9DBQ2 mus musculu
40	610.5	5.6	326	4	Q9H308	Q9H308 homo sapien
41	549.5	5.0	632	4	Q96N10	Q96N10 homo sapien
42	540	4.9	1105	10	Q23479	Q23479 arabidopsis
43	516	4.7	234	5	Q9VM05	Q9VM05 drosophila
44	429	3.9	1781	5	O17758	O17758 caenorhabd
45	390	3.6	1907	4	O15017	O15017 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q9B229 PRELIMINARY; PRT; 1539 AA.  
AC Q9B229; Q9UPU4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE BA15N3.2.1 (KIAA1058 protein) (Fragment).  
GN BA15N3.2.1 (KIAA1058 protein) (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith M.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 6-1539 FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Iehikawa K., Hirosewa M., Miyajima N., Tanaka A., Kozani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";  
RT DNA Rec. 6:197-205 (1999).  
DR EMBL; AL161420; CAC27814.1; -;  
DR EMBL; AB028981; BAA83010.1; -;  
FT NON TER 1 1  
SQ SEQUENCE 1539 AA; 176392 MW; 354F93C1120EBE71 CRC64;  
Query Match 73.2%; Score 8002; DB 4; Length 1539;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Cy 531 TLFKDSGNLDRKARFSAIYRQDSNKLKLLADFRKPEKRAKLPVILGNLDTTD 590  
Db 1 TLFKDSGNLDRKARFSAIYRQDSNKLKLLADFRKPEKRAKLPVILGNLDTTD 60

```

QY 591 NVSSDPNNVNSSYIPTRKQFETCSKPTTFEVEBEFVPCIPKHTQPTIYTNHLVYVPKYL 650
DB 61 NVSSDPNNVNSSYIPTRKQFETCSKPTTFEVEBEFVPCIPKHTQPTIYTNHLVYVPKYL 120
QY 651 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTRSAFAVLLHHQNPBF 710
DB 121 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTRSAFAVLLHHQNPBF 180
QY 711 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSWLPILKDRVVT 770
DB 181 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSWLPILKDRVVT 240
QY 771 SEOHIPVSNLPSGYLGYOELGMRHYGEIKWVDGKPLKISTHVSVTYTDQHLN 830
DB 241 SEOHIPVSNLPSGYLGYOELGMRHYGEIKWVDGKPLKISTHVSVTYTDQHLN 300
QY 831 PFQYCKTESGAQALGNELVKYKSLHAMEGHVMAFLPTILNQLFRVLTTRATQEEVAVN 890
DB 301 PFQYCKTESGAQALGNELVKYKSLHAMEGHVMAFLPTILNQLFRVLTTRATQEEVAVN 360
QY 891 VTRVLIHVVAQCHEEGLSHLSYVYKAYKAEPVYASEYKTVHEBELTKSMITILKESADF 950
DB 361 VTRVLIHVVAQCHEEGLSHLSYVYKAYKAEPVYASEYKTVHEBELTKSMITILKESADF 420
QY 951 LTSNKLKYSWFFDVILKSMQHLIENSKVKLLRQRPASVYHNAVETVVMMLMPHTQ 1010
DB 421 LTSNKLKYSWFFDVILKSMQHLIENSKVKLLRQRPASVYHNAVETVVMMLMPHTQ 480
QY 1011 KRDNPEASKANANSLAVFIPKRCFTFMDRGFVFKQINNYSICFAPGDPKTLPEYKEFLR 1070
DB 481 KRDNPEASKANANSLAVFIPKRCFTFMDRGFVFKQINNYSICFAPGDPKTLPEYKEFLR 540
QY 1071 VVCNHEHYIPLNMPFEGKRIQRYODLQDYSLTDEPCRNHVLGLLREVTALQOER 1130
DB 541 VVCNHEHYIPLNMPFEGKRIQRYODLQDYSLTDEPCRNHVLGLLREVTALQOER 600
QY 1131 EYRLAISVYKULLIHSFDDRYASRSHOARITLPLFGILLIENVQINVDVSPFV 1190
DB 601 EYRLAISVYKULLIHSFDDRYASRSHOARITLPLFGILLIENVQINVDVSPFV 660
QY 1191 NAGMTYKDESLALPAVNPVLPQKSTLNSLHKOLLGAIISGIASPYTTSTPNINSVNA 1250
DB 661 NAGMTYKDESLALPAVNPVLPQKSTLNSLHKOLLGAIISGIASPYTTSTPNINSVNA 720
QY 1251 DSRGSLISTDSGNSLPERNSEKSNLSLDRKQOSSTLGNVVRCDKLDQSEIKSLMCEFLYI 1310
DB 721 DSRGSLISTDSGNSLPERNSEKSNLSLDRKQOSSTLGNVVRCDKLDQSEIKSLMCEFLYI 780
QY 1311 LKSMSDALFTYNNKASTSELMDFITISEVCLHQFQYMGKRYIARTGMHARLQOLGSLD 1370
DB 781 LKSMSDALFTYNNKASTSELMDFITISEVCLHQFQYMGKRYIARTGMHARLQOLGSLD 840
QY 1371 NSLTFNHSYGHSDADVLHOSLLEANITATEVCLTALDTLSLFTLAFNQLADHGNPLMK 1430
DB 841 NSLTFNHSYGHSDADVLHOSLLEANITATEVCLTALDTLSLFTLAFNQLADHGNPLMK 900
QY 1431 KVPDYVLCLOKQOSSTALKNVFTALRSIYKPSSTFYGRADMCALCEILKCNSTL 1490
DB 901 KVPDYVLCLOKQOSSTALKNVFTALRSIYKPSSTFYGRADMCALCEILKCNSTL 960
QY 1491 SSIRTEASOLLYFLMNNFDTYKKSFPVTHLOVIVSGLADVVGIGTRFOQSLI 1550
DB 961 SSIRTEASOLLYFLMNNFDTYKKSFPVTHLOVIVSGLADVVGIGTRFOQSLI 1020
QY 1551 NNCANSRLIKITSSSDYKDLTKRIYTLAMATQKHEHNDPEMLVDLOYSLAKSYAST 1610
DB 1021 NNCANSRLIKITSSSDYKDLTKRIYTLAMATQKHEHNDPEMLVDLOYSLAKSYAST 1080
QY 1611 PELRKTWLSMARIHVKGNDLSBAANCYHVTALVAEYITRKAQVMEPPLPHSHSACL 1670
DB 1081 PELRKTWLSMARIHVKGNDLSBAANCYHVTALVAEYITRKAQVMEPPLPHSHSACL 1140

```

```

QY 1671 RRSRGVROGCTAARVITPNIDEASMEVGMQDVHFNEDVLMELLEQCADGLMKAER 1730
DB 1141 RRSRGVROGCTAARVITPNIDEASMEVGMQDVHFNEDVLMELLEQCADGLMKAER 1200
QY 1731 YELADIYKLIPIYEKRDPERLALYDTLHRAVSKYTEVMHSGRLLGTYFRVAFQO 1790
DB 1201 YELADIYKLIPIYEKRDPERLALYDTLHRAVSKYTEVMHSGRLLGTYFRVAFQO 1260
QY 1791 AAQOFTSETPDEGFEDDEKEXIYKPKLTPISISORLLKLYSDPFGSENVKMID 1850
DB 1261 AAQOFTSETPDEGFEDDEKEXIYKPKLTPISISORLLKLYSDPFGSENVKMID 1320
QY 1851 SGKVPKOLDSKYAYIOVTHVIPFDEKELQERKTEPERSHNIIRRFMEEMPTQTKQG 1910
DB 1321 SGKVPKOLDSKYAYIOVTHVIPFDEKELQERKTEPERSHNIIRRFMEEMPTQTKQG 1380
QY 1911 GVEEOCKRTITLTAHCPRYKRIIPWYQHTDLNPIEVAIDENSKYAEIROLCSAE 1970
DB 1381 GVEEOCKRTITLTAHCPRYKRIIPWYQHTDLNPIEVAIDENSKYAEIROLCSAE 1440
QY 1971 VDMIKLOLQKQSVSVQVNAAGPLAARAFLODPTNTKRYPDNKVKLLKEVFRQFVACGQ 2030
DB 1441 VDMIKLOLQKQSVSVQVNAAGPLAARAFLODPTNTKRYPDNKVKLLKEVFRQFVACGQ 1500
QY 2031 LAVNERLIKEDQLEJOEEMKANYREMAKELSEIMEQI 2068
DB 1501 LAVNERLIKEDQLEJOEEMKANYREMAKELSEIMEQI 1538

```

```

RESULT 2
Q9B227
ID Q9B227; PRELIMINARY; PRT; 909 AA.
AC Q9B227;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BA155N3.2.2 (KIAA1058) (Fragment).
GN BA155N3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL161420; CAC27816.1; -.
FT NON_TER 1
FT TER 909
SQ SEQUENCE 909 AA; 103445 MW; 77166F77937339F3 CRC64;

```

```

Query Match 42.1%; Score 4605.5; DB 4; Length 909;
Best Local Similarity 97.5%; Pred. No. 2,5e-285;
Matches 886; Conservative 0; Mismatches 0; Indels 23; Gaps 1;

```

```

QY 531 TLFKASGNDLNKAFSAIYRQDSNKLNDMDKLADFRKPEKAKLPIVIGNDITD 590
DB 1 TLFKASGNDLNKAFSAIYRQDSNKLNDMDKLADFRKPEKAKLPIVIGNDITD 60
QY 591 NVSSDPNNVNSSYIPTRKQFETCSKPTTFEVEBEFVPCIPKHTQPTIYTNHLVYVPKYL 650
DB 61 NVSSDPNNVNSSYIPTRKQFETCSKPTTFEVEBEFVPCIPKHTQPTIYTNHLVYVPKYL 120
QY 651 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTRSAFAVLLHHQNPBF 710
DB 121 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTRSAFAVLLHHQNPBF 180
QY 711 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSWLPILKDRVVT 770
DB 181 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSWLPILKDRVVT 240
QY 771 SEOHIPVSNLPSGYLGYOELGMRHYGEIKWVDGKPLKISTHVSVTYTDQHLN 830

```

```

Db 241 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWVGKPLKISTHLYSTVYTDQOHLN 300
Qy 831 FFQYCKTESGAQALGNELVYKLSLHMEGHVMIAPLPTLNQLFVLTATOEBAVN 890
Db 301 FFQYCKTESGAQALGNELVYKLSLHMEGHVMIAPLPTLNQLFVLTATOEBAVN 360
Qy 891 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 950
Db 361 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 420
Qy 951 LTSNKLKYSWFFPDVLKSMQOHLIENSKYKLRNORFPASYHNAVEVVMMLPHITQ 1010
Db 421 LTSNKLKYSWFFPDVLKSMQOHLIENSKYKLRNORFPASYHNAVEVVMMLPHITQ 480
Qy 1011 KFRDPEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDKPTLFEYKEFLR 1070
Db 481 KFRDPEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDKPTLFEYKEFLR 540
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRVGTALQEFR 1130
Db 541 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRVGTALQEFR 600
Qy 1131 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 1190
Db 601 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 660
Qy 1191 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTSTPNINSVNA 1250
Db 661 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTSTPNINSVNA 720
Qy 1251 DSRGSLISTDSGNSLPERNSEKSNLDRHQOQSTLGNSVVRCDKLDQSEIKSLMCFLYI 1310
Db 721 DSRGSLISTDSGNSLPERNSEKSNLDRHQOQSTLGNSVVRCDKLDQSEIKSLMCFLYI 780
Qy 1311 LKSMDDALFTYWNKASTSELMDFPTISEVCLHOPQYMKRYIARNOQELGPIVHDRSQ 840
Db 781 LKSMDDALFTYWNKASTSELMDFPTISEVCLHOPQYMKRYIARNOQELGPIVHDRSQ 840
Qy 1355 -----RTGMMHARLQOGLSDNSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDT 1407
Db 841 TLPVGRNRGTGMHARLQOGLSDNSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDT 900
Qy 1408 LSLFTLAFK 1416
Db 901 LSLFTLAFK 909

RESULT 3
Q9B228 PRELIMINARY; PRT; 878 AA.
ID Q9B228;
AC Q9B228;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BA155N3.2.3 (KIAA1058) (Fragment).
GN BA155N3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, AL161420; CAC27815.1; -.
FT NON_TER 1
FT NON_TER 878
SQ SEQUENCE 878 AA; 99777 MW; C9475D4FE68B7231 CRC64;

Query Match 41.1%; Score 4491; DB 4; Length 878;
Beet Local Similarity 97.4%; Pred. No. 4, 9e-278;
Matches 863; Conservative 6; Mismatches 9; Indels 8; Gaps 1;

```

```

Qy 531 TLPKASGNLDKNARFSAIYRQDSNKLSDMDMLKLLADFRKREKAKLPIVLGNLDTID 590
Db 1 TLPKASGNLDKNARFSAIYRQDSNKLSDMDMLKLLADFRKREKAKLPIVLGNLDTID 60
Qy 591 NVSSDFPNYVNSYIPTQEFCTCKPTTFEVEEVPICIPGHTOPYTITMNLVYPRYL 650
Db 61 NVSSDFPNYVNSYIPTQEFCTCKPTTFEVEEVPICIPGHTOPYTITMNLVYPRYL 120
Qy 651 KYDSOKSFAKARNIACIEFPDSDEBQPLKCIYGRGSPVFRSAPAAVLHHQNEF 710
Db 121 KYDSOKSFAKARNIACIEFPDSDEBQPLKCIYGRGSPVFRSAPAAVLHHQNEF 180
Qy 711 YDEIKIELPTQHEKHLLTFPHVSCNNSKSGTKKRDVETQVGYMPLLDKGRVYT 770
Db 181 YDEIKIELPTQHEKHLLTFPHVSCNNSKSGTKKRDVETQVGYMPLLDKGRVYT 240
Qy 771 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWVGKPLKISTHLYSTVYTDQOHLN 830
Db 241 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWVGKPLKISTHLYSTVYTDQOHLN 300
Qy 831 FFQYCKTESGAQALGNELVYKLSLHMEGHVMIAPLPTLNQLFVLTATOEBAVN 890
Db 301 FFQYCKTESGAQALGNELVYKLSLHMEGHVMIAPLPTLNQLFVLTATOEBAVN 360
Qy 891 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 950
Db 361 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 420
Qy 951 LTSNKLKYSWFFPDVLKSMQOHLIENSKYKLRNORFPASYHNAVEVVMMLPHITQ 1010
Db 421 LTSNKLKYSWFFPDVLKSMQOHLIENSKYKLRNORFPASYHNAVEVVMMLPHITQ 480
Qy 1011 KFRDPEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDKPTLFEYKEFLR 1070
Db 481 KFRDPEASKNANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDKPTLFEYKEFLR 540
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRVGTALQEFR 1130
Db 541 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRVGTALQEFR 600
Qy 1131 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 1190
Db 601 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 660
Qy 1191 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTSTPNINSVNA 1250
Db 661 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTSTPNINSVNA 720
Qy 1251 DSRGSLISTDSGNSLPERNSEKSNLDRHQOQSTLGNSVVRCDKLDQSEIKSLMCFLYI 1310
Db 721 DSRGSLISTDSGNSLPERNSEKSNLDRHQOQSTLGNSVVRCDKLDQSEIKSLMCFLYI 780
Qy 1311 LKSMDDALFTYWNKASTSELMDFPTISEVCLHOPQYMKRYIARNOQELGPIVHDRSQ 840
Db 781 LKSMDDALFTYWNKASTSELMDFPTISEVCLHOPQYMKRYIARNOQELGPIVHDRSQ 840
Qy 1371 NSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDTLSLFTLAFK 1416
Db 834 -GISVDNGYGSHDADVLHOSLEANIATEVCLTALDTLSLFTLAFK 878

RESULT 4
Q9B226 PRELIMINARY; PRT; 724 AA.
ID Q9B226;
AC Q9B226;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BA155N3.2.4 (KIAA1058) (Fragment).
GN BA155N3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith M.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AL61420; CAC27817.1; -  
 FT NON TER  
 SO SEQUENCE 724 AA; 82617 MW; 7E2B345FE2FE63D6 CRC64;  
 Query Match 33.8%; Score 3693; DB 4; Length 724;  
 Best Local Similarity 97.1%; Pred. No. 3.3e-227;  
 Matches 708; Conservative 1; Mismatches 6; Indels 14; Gaps 1;  
 QY 531 TLEKDSAGNLDKNAKARSATYRQDSNKLSDMDMLKADLPRKPEKMAKLPEVILGNLDTITD 590  
 DB 1 TLEKDSAGNLDKNAKARSATYRQDSNKLSDMDMLKADLPRKPEKMAKLPEVILGNLDTITD 60  
 QY 591 NVSSDPNNVNSSYITPKQETCSKTPITEVEEFPVPCIPKDTOPYTIYTNMLYVYPKYL 650  
 DB 61 NVSSDPNNVNSSYITPKQETCSKTPITEVEEFPVPCIPKDTOPYTIYTNMLYVYPKYL 120  
 QY 651 KYDSOKSPAKARVIAICIEFKDSDEDSQPLKCTYGRPGGPVTRSAFAVILHHONPER 710  
 DB 121 KYDSOKSPAKARVIAICIEFKDSDEDSQPLKCTYGRPGGPVTRSAFAVILHHONPER 180  
 QY 711 YDEIKTELPTQHEKHLHLTFFHVSQDSSKSTKKRDVETQVGSMLPLKDGRTVT 770  
 DB 181 YDEIKTELPTQHEKHLHLTFFHVSQDSSKSTKKRDVETQVGSMLPLKDGRTVT 240  
 QY 771 SEGHIVSANTLPGCYGQYELGKGRHYGPEIKVNDGSKPLKSTHVLSTVYTQDQHLN 830  
 DB 241 SEGHIVSANTLPGCYGQYELGKGRHYGPEIKVNDGSKPLKSTHVLSTVYTQDQHLN 300  
 QY 831 FPOYCKTSSGALGNELVKYKLSHAMEGHVIAFLPTILNQLTRVLTTRAQSEAVN 890  
 DB 301 FPOYCKTSSGALGNELVKYKLSHAMEGHVIAFLPTILNQLTRVLTTRAQSEAVN 360  
 QY 891 VTRVLIHVVAQCHEGLESHTSRYSYVYAAKABEYVASEKTVHEBLTKSMITILKPSADP 950  
 DB 361 VTRVLIHVVAQCHEGLESHTSRYSYVYAAKABEYVASEKTVHEBLTKSMITILKPSADP 420  
 QY 951 LTSNKLKTSWFFDYVILKSMAGHLIENSKVKLLNQRPPASTYHAAVEVYVNMAMPHITQ 1010  
 DB 421 LTSNKLKTSWFFDYVILKSMAGHLIENSKVKLLNQRPPASTYHAAVEVYVNMAMPHITQ 480  
 QY 1011 KRPDNEASAKNAHSLAVFKRCFTMDRGFVKQNNNTYISGAPGDPKTLFEYKEEFLR 1070  
 DB 481 KRPDNEASAKNAHSLAVFKRCFTMDRGFVKQNNNTYISGAPGDPKTLFEYKEEFLR 540  
 QY 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLQDLSLTDFCRNHFVGLLREVGTAQOEFR 1130  
 DB 541 VVCNHEHYIPLNLPMPFGKRIQRYODLQDLSLTDFCRNHFVGLLREVGTAQOEFR 600  
 QY 1131 EYRLIAISYVKNLLIHSFDDRYASRSHQRIATLPLFGILLIENVORINRVDSPPV 1190  
 DB 601 EYRLIAISYVKNLLIHSFDDRYASRSHQRIATLPLFGILLIENVORINRVDSPPV 660  
 QY 1191 NAGMTYKDSIALPAVNPLVTPQKSTLDSLHKDLIGIASYTTSTPINSVRNA 1250  
 DB 661 NAGMTYKDSIALPAVNPLVTPQKSTLDSLHKDLIGIASYTTSTPINSVRNA 706  
 QY 1251 DSRGSLIST 1259  
 DB 707 PCSGGLST 715  
 RESULT 5  
 ID Q9VMO4 PRELIMINARY; PRT; 1984 AA.  
 AC Q9VMO4;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (T-EMBLrel. 13, Last annotation update)

DE CG6630 protein.  
 GN CG6630.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Galburt W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mantei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter R., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL, AB003617; AAF5824.1; -  
 DR FlyBase; FBgn0031921; CG6630.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PSS0003; PH DOMAIN; 1.  
 SO SEQUENCE 1984 AA; 224875 MW; 55441A434022712F CRC64;  
 Query Match 30.2%; Score 3305.5; DB 5; Length 1984;  
 Best Local Similarity 37.0%; Pred. No. 9.6e-202;  
 Matches 767; Conservative 343; Mismatches 615; Indels 347; Gaps 49;  
 QY 45 KPKLIEPLDYENVYQKQTQILNDCLREMLLFPPYDDPQTALIRQGRYICSTV----- 97  
 DB 5 KPLVVEPIDFEFAFKAKNTVITQNDPQRRLILYPPDDVSEIIMPKRQRTNMAVSADRPDP 64  
 QY 98 -----PAAEEEAQ----- 97  
 DB 65 NEAIVCPHAGSPMITNGNGHGSQVSRQSGISQNGSHHNGNGHNTSSSSSSSLNNSNGHQ 124  
 QY 98 -----PAAEEEAQ----- 106  
 DB 125 LSRKSSQCSNGSSSQKDSYSALSSITLRSNLAQPEVDEFAADGNADLVQGPQHSR 184

QY 107 ---SLPTECIKTNSDMHLVNYKYEDYSGEERQJPNKVVLDKLPVHYVEDEVDKDE 163  
 DB 185 AECTFRQALYTYPAKHLIHYKYNAGCNCHDLP-16PAEBELVEEITIDADQDRID 243  
 QY 164 DAASIGSQKGTITKGMLYKGMNSA----ISVTRSPFRHPLIQLQDGSYKEPEFLKD 219  
 DB 244 EOMT-RSQADITTYKQGYLLKGPDSADRMFANIGKSPFRHICYLRQETIDGYILIEHMD 302  
 QY 220 -LOKEPKSIGFLGYVSPFRNNKVRPAFELKM-ODKSSVLLADSEVEMETITLTK 277  
 DB 303 EKQSAKATIVMDFCTDV-VQNPKGGRFCFELRMTHGKSTLLAENEDPFQWLSKISS 361  
 QY 278 ILQJNPEAMOKRND-----SHDEQSKLE-----GSGSLDSTYL-PELAKSA 322  
 DB 362 VLAQN--RAOEKRNALSERHPSIGSNPSPQLQPRAMPDPFTGTLKGLDQSLHPQLMKYG 419  
 QY 323 REAEIKL---KSESVKLP--YLDPDAQLDSSAEPVYKSPBEKFGRIYLYKCDLSN 377  
 DB 420 RETHSIALARREQRRLPACYQSPVK-----GSGSDNEQYREHFGRTLLTCHLRER 474  
 QY 378 LOCCVAENEB---GPTTNEPPEFVTLSPFDIKYNNKISADPHVDLNFESVRQMIATTS-P 433  
 DB 475 LQ-CIPODESSAGVBOQVEPYITSLALPDAAKANKLSENFYFNNEQPAQDLPTPTVP 533  
 QY 434 ALMNGSGPEQALRGILHEAMQY-----KQIFSYTCBP 471  
 DB 534 SSVACQGVPRKA-BGDERSTACQAPHSLPDGVSAELLSNRQOQQLRGCLSLTAPHA 592  
 QY 472 DIFLVAIEKVLQSGITHCABPYMKSDDSKYAQKYLKNAKQACQRLGQYRMPFMAART 531  
 DB 593 DYLVAIREKILQSGIAQVTEPYLXAGKDPKLGQKYKAAKSYAQHIGHYRPPFMAAR 652  
 QY 532 LFKDASGNLD---KNARSAIYRDOONSLNDMLKLABPKPEKAKLVIIIGNDI 587  
 DB 653 LFKQYSHEDVDPKKBEFSPPIYRQETPKLDEBELIKLVDRKEBKSKLITIGSLKM 712  
 QY 588 TIDNVSDEPNYVNSYIPTKQETCSKPTIFEEVEFVPCIPKHTOPYITYNHLYVP 647  
 DB 713 QMQLDQTTPCGLSKSLAPLSTFSPSSKQSPTEVAEFQSGSEBRADHPTSCNHLVYP 772  
 QY 648 KTLKIDSQSPAKANIALICIEFKSDDEDSOPLKCTYRPGCPVFTSAPAAVTHHON 707  
 DB 773 LSLQPSOQLFRRARITVIVELRQDGEYSKPLKCIYGRPGODLLVSGIACPVLHNVT 832  
 QY 708 PEFYBEIKI ELPTOLHEKHLITLTPHVSCDNSKSGSTKRP---VVEQVQYSMLPLK 764  
 DB 833 PTWBEIKIKRLPLGLFPEHHLIFSIFYHSCNLS---KRDANAFAEPPIGIALMLPLQ 887  
 QY 765 DGRVVTSEQHIFVSANLPSGYLGYOELGMR--HYGEIKWYDGGKPLIKISTHVLSTVY 822  
 DB 888 KNRICIEBQQLVVAATLPVGYLSIQPLGKGCQNGCPDICWDNGRNLYTVGLARDSTYL 947  
 QY 823 TODOHLANFQYQKQT---ESGAQALGNELVYLYKSLPMEGHVMIAFLPTILNQLFRV 878  
 DB 948 TADOHLANFPGHCEBRLBEGKTGAVPAETETCKILKAHAIDMKSLINLYPTVLNELFTL 1007  
 QY 879 LIRAQOEBAVAVVTYIIHVVAQCHEEGLESH-LBSYVYAKAKBEPVASE---YKTVAE 934  
 DB 1008 LVHTQSEEGVAVIRITINIHLISDQAKRSOLLGAYVAVFPA-PYYSQCARQARTVHG 1066  
 QY 935 BLTKSMTTLKPS-ADPLTSNKLKYSNFPFVLIKSMACHLJENSKVLLNNORFPASY 993  
 DB 1067 ELCHRLHLYLANSNPFLVKNFMRYSSTFFPLIISKMAQHLLALGIRIMANERFPKXY 1126  
 QY 994 HHAVETVVMMLPHITQKFRNDPEASKANHSLAVFIKQCTFPMRGRVFKQINNYISCF 1053  
 DB 1127 GDRVEGLIALMPYITRREDISEEETHLNRSLAKRVRCCLTYMRGFRILIRCMGEF 1186  
 QY 1054 ARGDPTLSEYKFEPLRVVCMNEHYIPLNLPMPFGKGRQRYODQLDSDLDEFCRNF 1113  
 DB 1187 SGNPRILHEYKFNFLQEBICQEHYVPLN--PVLNPKRPPENMHQHTLSEQFRCOF 1244  
 QY 1114 LVGLLREVGTALQEFREVLALIAISVLKNLIIKHSFDEYASRSHQARITATLYLPLGLL 1173

DB 1245 LSGILLQELKSLNEVGHVRHALLGIFKDLAKHELDNRXQOKOLSHIALLYPMLGV 1304  
 QY 1174 IENQRIVRDVPSPFNAGTVDK---SLALAVNPLVYPOKSTLND----- 1220  
 DB 1305 MDNHRIDDLSESGACTPNGVTVADSAHYKRLSCSSGVYFSKOSTGSLTSPRKNR 1364  
 QY 1221 -SLKHDLGALISGIALPYTST-----PINSVNRADSGSLISTDSGNSL 1265  
 DB 1365 LTHMCDQ-----ASPRYSVHKHKNHYLAAIAGQPIISGINSLSLUS---NNDSGHSQ 1414  
 QY 1266 PE-----RNSKNSL-DKQOSSTLGNV--RCDKLDSEIKSLMCTLYLYKMSD 1316  
 DB 1445 DTTTIGAYTNDDTVALNGHNRSVSYTHAOILRCDKFSSVESKDLILGLFLIKHLSQ 1474  
 QY 1317 DALFTYMKASTSELMDFITSEVCLHQFOYMGKRYI-----ARTGMHARIQ 1365  
 DB 1475 EOMIGMWCNSESLEQLSLDCLQFRYVGKSVVITTEGRLAKANTLPARTOP 1534  
 QY 1366 LGSIDN-----SLTFNHSYGH--SPADVLHO---SLBANIAVECLTALDLSLFT 1412  
 DB 1535 PTGLNSEQBOQPSGCTINQREHLBEDIDTLANQALAYESNLATBGMIIIDCLGLYV 1594  
 QY 1413 LAFKNQLIADHGNPLMKVFDVYLCFLQKQSELTALKNVFTALRSLYKEPSTFYSGRA 1472  
 DB 1595 LQFR-QLLAD---SLVLPKVARVYLRFLQJQCSERLSGHPAALRAPINNYAVALFKGNA 1650  
 QY 1473 DMCAALCEILKCNKSLSTIRTEASQLYFLMNNPNYOKKS FVRHLOVITSVSGLI 1532  
 DB 1651 MLCQMVEVELLKACDSRLVEIRHSCAVLYLLMKSNEPFSGRKALTRVHLOVITSVSOMI 1710  
 QY 1533 ADVVIGITRFOQSLSIINNCANDRLIKHTSFSSDVCDLTKIRTVLMATAOMKEHND 1592  
 DB 1711 GNVIGLNNARKQBSLSIINSTANSBDMKMGKGFMEVVDLTRRTVLMATQOAHMMD 1770  
 QY 1593 PEMLVLDQYSLAKSYASTPELRTKWLDSMAIIVKNGDLSAAMCYVHTALVAEYLTRK 1652  
 DB 1771 PERLLEIQYSLANSYASTPELRHTWLVYMAHNGNGLSBAACHLIALMCEYL--- 1827  
 QY 1653 EAVQMEPRLPHSHSACLRNRG-VFRQGTARVITPNID-BEASNMEDVGMQDVFN 1710  
 DB 1828 -----RLKGGCTLSWSSTAIGKSTNPLDQGIKADAGDDSOYT 1868  
 QY 1711 EDVIMLEBOCADGLMKAREYELIADYKLIIPYEKARDPERLAHLYDLRAYSKYE 1770  
 DB 1869 EOMLEFOKLQADPLDRERFECGELYKILPMEYRERSQDLACHENHLOANUK-- 1926  
 QY 1771 VMHSGRLLGTFRVAFFGQAQYQFTDSETFVEGFEDDEKEXIYKPLTLPSEISQ 1830  
 DB 1927 -----VENYFEDHAIBRVYKPEKLTLSLSISE 1954  
 QY 1831 RLULKYSDFGSBNVXMIQDSGKYNPKDLSK 1862  
 DB 1955 RLAKQYKEKFGADVVKMIMDS---SPVFLDWK 1983  
 RESULT 6  
 Q63603 ID ID 063603 PRELIMINARY; PRT, 738 AA.  
 AC 063603;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE TRG protein (Fragment).  
 GN TRG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN NCB1\_TaxId=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISCHER;  
 RA Pianeze L., Porcellini A., Avvedimento V.E., D'Esposito F.,

RA Felleiello A., Monticelli A., Mueti A.M., Tortora G., Varrone S.,  
 RA Cocozza S.,  
 RA Life Sci. Adv. (Mol. Biol.) 13:75-83 (1994).  
 DR EMBL: X68101; CA48220.1; -  
 FT NON\_TER 1 1  
 SQ SEQUENCE 738 AA; 84857 MW; 5E354ED2A34F977 CRC64;

Query Match 26.4%; Score 2888; DB 11; Length 738;  
 Best Local Similarity 87.1%; Pred. No. 9e-176;  
 Matches 573; Conservative 23; Mismatches 36; Indels 26; Gaps 7;

QY 1424 GNNPLMKKVPDYVLCFQKQKSTALKNVFTALRSIIYFPSTFYGRADMALCYEIL 1483  
 DB 5 GHSPLMKKVPDYVLCFQKQKSEMAKNVFTALRSIIYFPSTFYGRADMALCYEIL 64  
 QY 1484 KCCNSLSSIRTEASQLVFLMNNPDYCKGSFVYTHLOVITISVQLADVVGIGTGF 1543  
 DB 65 KCCNSLSSIRTEASQLVFLMNNPDYCKGSFVYTHLOVITISVQLADVVGIGTGF 124  
 QY 1544 QQSLSIINNCANDRLIKHTSFSSDYKDLTKRIRYLMATAOMKEHNDPEMLVDLYSL 1603  
 DB 125 QQSLSIINNCANDRLIKHTSFSSDYKDLTKRIRYLMATAOMKEHNDPEMLVDLYSL 184  
 QY 1604 ANSYASTPELRKTWLDSPARIHVKNCDLSEANACVYVHTALVAEYLRKE--AVQWEP 1660  
 DB 185 ANSYASTPELRKTWLDSPARIHVKNCDLSEANACVYVHTALVAEYLRKEADLALQREBP 244  
 QY 1661 LPHSHSACLRRSGVFRGCTAFRVTIPNIDEEASMMEDVGMVHNEDVLMELLBO 1720  
 DB 245 VFPYSHTSCKRSRGMFPGCTAFRVTIPNIDEEASMMEDVGMVHNEDVLMELLBO 304  
 QY 1721 CADGLKKAERYE--LIADYKLIPIYERKRDEFLAHYDTHLRAYSKYTEVHSGRL 1778  
 DB 305 CADGLKKAERYE--LIADYKLIPIYERKRDEFLAHYDTHLRAYSKYTEVHSGRL 359  
 QY 1779 LGTYFRV--AFGQAAQYOFDSETEVEGFEDDEKEYIYKPKLTPSEISQRLIKY 1836  
 DB 360 AGSWDLPLGOLFQ-----GFEDDEKEYIYKPKLTPSEISQRLIKY 405  
 QY 1837 SDKFGSENVKMIODSGKVPKLDISKYAYIQTHTVHPFDEKELORKEPESHNRFP 1896  
 DB 406 SDKFGSENVKMIODSGKVPKLDISKYAYIQTHTVHPFDEKELORKEPESHNRFP 465  
 QY 1897 MEMPRTQTGKRGQVEEOCKRTILTAIHCFPYVKRIIPWYQHHTDLPLEVAIDENS 1956  
 DB 466 MEMPRTQTGKRGQVEEOCKRTILTAIHCFPYVKRIIPWYQHHTDLPLEVAIDENS 525  
 QY 1957 KKYAELPOLCSSAEVDMIKLQKLQGSVSQVNAAGPLAARAFLDNTNKRYPDNKVKLL 2016  
 DB 526 KKYAELPOLCSSAEVDMIKLQKLQGSVSQVNAAGPLAARAFLDNTNKRYPDNKVKLL 585  
 QY 2017 KEVFRQFVACGQALVNERLIKEPDLEYOEEMKAYREMAKELSIIMEQICPLEEK 2074  
 DB 586 KEVFRQFVACGQALVNERLIKEPDLEYOEEMKAYREMAKELSIIMEQICPLEEK 643

## RESULT 7

Q96MN3 PRELIMINARY; PRT; 881 AA.

AC Q96MN3; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE CDNA FLJ32122 f19, clone PEBL1000144, moderately similar to Trg.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 (1) SEQUENCE FROM N.A.  
 RA Ieshbaeti T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Houta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,

RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hito Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Magatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Naganari K., Masuno Y., Nagai K., Isogai T.,  
 RT "NEO human cDNA sequencing project".  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK056884; BAB71253.1; -  
 SQ SEQUENCE 881 AA; 100704 MW; 351B6D432ADBA2A CRC64;

Query Match 25.1%; Score 2739.5; DB 4; Length 881;  
 Best Local Similarity 61.2%; Pred. No. 3.7e-166;  
 Matches 539; Conservative 128; Mismatches 140; Indels 73; Gaps 8;

QY 1240 STPNINSVNMDSRSLSTDSGNSLPRNSKSNLSLKHQSSSTLGNVRCODLDDSE 1299  
 DB 33 SFQNHGIRKEDSRSL--PEGATCFPQNGTNGT---KQST-RSSVQYNLDDYE 86  
 QY 1300 IKSILMCELYILKSSDDELFTYNNKASTSEIMDFITSEVCLHQFOYMGKRYIA----- 1354  
 DB 87 IRSILMCELYIYKMSIEDTLTYNNKVSPOELINILILEVCLPHFRYMGKRIARVDA 146  
 QY 1355 -----RTGMHARLQOLGSLDNLSTNNHYGSDADVLHOSLLBA 1394  
 DB 147 WLSKHFQIDRSQTPALNRSGVQARLOHLSLSESTFNNHSETTEADI FHOALDEG 206  
 QY 1395 NIAETVCLTALDTSFLTAFNOLLADHGNPLMKKVPDYVLCFQKQKSTALKNVFT 1454  
 DB 207 NIAETVSLVLDTSISFTQCFKTLNNDGNPLMKKVPDYVLCFQKQKSEVSLKHPA 266  
 QY 1455 ALRSLIYFPSTFYGRADMALCYEILKCCNSLSSIRTEASQLVFLMNNPDYCKG 1514  
 DB 267 SLRAFISKFPSPAFGRVNMCAFCYEVLCCKTSTISSTRBASALVFLMNNPEYTKR 326  
 QY 1515 KSFVTHLOVITISVQLADVVGIGTGFQOQSLIINNCANDRLIKHTSFSSDYKDLTK 1574  
 DB 327 KTFELRTHLOVITISVQLADVVALSGSFRQESLFINNFANSDREPKATAPEAKYDLTK 386  
 QY 1575 RIRYLMATAOMKEHNDPEMLVDLYSLASVASTPELRKTWLDSPARIHVKNCDLSEA 1634  
 DB 387 RIRYLMATAOMKEHNDPEMLVDLYSLASVASTPELRKTWLDSPARIHVKNCDLSEA 446  
 QY 1635 AMCYVHTALVAEYLRKEAVQWEPPLPHSHSACLRRSGVFRGCTAFRVTIPNIDE 1694  
 DB 447 AMCYVHTALVAEYLRKE-----LFPNGSARFKITPINIDE 483  
 QY 1695 EASMMEDVGMVHNEDVLMELLBOCADGLKKAERYELIADYKLIPIYERKRDEFL 1754  
 DB 484 EGAMKEDAGMNDVHSEVLLLELQCVDMKKAERYELIADYKLIPIYERKRDEFL 543  
 QY 1755 AHLVDTLRAYSKYTEVHSGRLIGTYFRVAFGQAAQYOFDSETEVEGFEDDEKE 1814  
 DB 544 TOVRTLHGATYKLELVNHTKRLGLGTFRAVFGQS-----FEEEDKE 589  
 QY 1815 YIYKPKLTPSEISQRLIKLQYSDKFGSENVKMIODSGKVPKLDISKYAYIQTHTVHP 1874  
 DB 590 YIYKPKLTPSEISQRLIKLQYSDKFGSENVKMIODSGKVPKLDISKYAYIQTHTVHP 649  
 QY 1875 FDEKELORKEPESHNRFPMEMPRTQTGKRGQVEEOCKRTILTAIHCFPYVKR 1934  
 DB 650 FDDKELTERKTEFERNHNISRFVEAPYTLGSKQKGCIEEOCKRTILTTNSFPYVKR 709  
 QY 1935 IPWYQHHTDLPLEVAIDENS KKYAELPOLCSSAEVDMIKLQKLQGSVSQVNAAGPLA 1994  
 DB 710 IPINCEQOINKPDIADDEIKDTAELOKLCSSIDVMDIOLQKLQGSVSQVNAAGPLA 769  
 QY 1995 YARAFLDNTNKRYPDNKVKLLKEVFRQFVACGQALVNERLIKEPDLEYOEEMKAYR 2054  
 DB 770 YARAFLDNTNKRYPDNKVKLLKEVFRQFVACGQALVNERLIKEPDLEYOEEMKAYR 829  
 QY 2055 EMKELSIIMEQICPLEEKSTVLP--NSLHFNALISGT 2091  
 DB 830 DMVXELSDIIEQI--IQEDTMHSPMNSNTLHVFCAISGT 867



```
RESULT 8
Q9BZ25 PRELIMINARY; PRT; 448 AA.
ID Q9BZ25;
AC Q9BZ25;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE BA155N3.1 (Novel protein similar to KIAA0694) (Fragment).
CN BA155N3.1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161420; CAC27818.1; -
DR InterPro; IPR001064; Crystal1in.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00003; PH DOMAIN; 1.
FT NON TER
SQ SEQUENCE 448 AA; 50843 MW; 1238F8B5FALA27D CRC64;

Query Match 20.0%; Score 2192; DB 4; Length 448;
Best Local Similarity 95.1%; Pred. No. 1.2e-131;
Matches 428; Conservative 6; Mismatches 12; Indels 4; Gaps 3;

QY 83 TAILRQGGYICSTVPKAKEEBAOSLFVTECTKTNSDMHLVNYKEDYSGSEFROLPNKY 142
DB 1 TAILRQGGYICSTVPKAKEEBAOSLFVTECTKTNSDMHLVNYKEDYSGSEFROLPNKY 60
QY 143 VLDLKLPAVAVVEDEVDKEDDAASIGSGGITTGHGMLYKGNMNSAISVTWRSFKRPF 202
DB 61 VLDLKLPAVAVVEDEVDKEDDAASIGSGGITTGHGMLYKGNMNSAISVTWRSFKRPF 120
QY 203 HLIQLGDSGYKEPELKD-LQKEPKGSIFLGPLYGVSPNNKVRPAPELKNQDSSYLL 260
DB 121 HLIQLGDSGYNLFYKDEKISKEPKGSIFLDSCMGV-VQNNKVRPAPELKNQDSSYLL 179
QY 261 AADSEVEMEWITTIINKILQINPEAMQKRGSDSHEDDSKLBSSGSGLDLSYLPBLAK 320
DB 180 AADSEVEMEWITTIINKILQINPEAMQKRGSDSHEDDSKLBSSGSGLDLSYLPBLAK 239
QY 321 SARAEAIKLKSESRVQLFYLDPAQKLDPSAEPYKSFEEKFKRIILVKCNDLSFNLQC 380
DB 240 SARAEAIKLKSESRVQLFYLDPAQKLDPSAEPYKSFEEKFKRIILVKCNDLSFNLQC 299
QY 381 CVAENBEGFTTNEPFTVTLSPDIKYNNKISADPFVDNLHFSVROMIATTSPALNNGSG 440
DB 300 CVAENBEGFTTNEPFTVTLSPDIKYNNKISADPFVDNLHFSVROMIATTSPALNNGSG 359
QY 441 PETGQALRGITLHEAMQYKPGIFSVTCBHPDILVARIEKYLQSGITHCAPYKSSNS 500
DB 360 PETGQALRGITLHEAMQYKPGIFSVTCBHPDILVARIEKYLQSGITHCAPYKSSNS 418
QY 501 SKVAQKVLNAKQACORLGOYRPFAMAR 530
DB 419 SKVAQKVLNAKQACORLGOYRPFAMAR 448

RESULT 9
Q9MX10 PRELIMINARY; PRT; 611 AA.
ID Q9MX10;
AC Q9MX10;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
```

```
DE BA12804.1.1 (Novel protein) (Fragment).
CN BA12804.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lawlor S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391280; CAD21444.1; -.
FT NON TER
SQ SEQUENCE 611 AA; 70252 MW; 1B0B6CB9AC2E1AC CRC64;

Query Match 19.9%; Score 2180; DB 4; Length 611;
Best Local Similarity 66.4%; Pred. No. 1.1e-110;
Matches 421; Conservative 88; Mismatches 83; Indels 42; Gaps 4;

QY 1463 PPSFTFEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLMRNPNPYTKSGSPVRTL 1522
DB 1 PPSAFTFKGRVNMCAFCYEVLLKCTSKISIRNEASALLYLMRNPFYTKRTFLRTL 60
QY 1523 QVILSVGLINDVYIGGTRFQOSLSTNNCANDRLIKHTSPSSDVYDLTKRIRTYLMA 1582
DB 61 QVILAVSQLIADVALSGGSRFQBSLFIINNPNANDRPKATAPFAEVZDLTKRIRTYLMA 120
QY 1583 TAOMKEHNDPEMLVDLOYSLAKSYASTPELRKTMDSMARHYVNGDLSAAMCYVHT 1642
DB 121 TAOMKEHNDPEMLVDLOYSLAKSYASTPELRKTMDSMARHYVNGDLSAAMCYVHT 180
QY 1643 ALVAEYLTRKEAVOWEPELPHSHASACLRSGGVFROGCTAFRVTIPNIDBSAMEDV 1702
DB 181 ALVAEFLHRKK-----LFRGCGAFKKTITENIDEGAMKEDA 217
QY 1703 GMQDYHFNEDVLMELBGCADGLWAKERYELIADIKLITFYEKRRPEKLAHLYDTLH 1762
DB 218 GMQDYHVEEVLLELBQCVDGLWAKERYELIISESKLIVBIYKRRPEKLTQVRYTLH 277
QY 1763 RAYSKVTEVMSGRRLGTFYRVAFFGOAAQYQFTSETDVEGFPEDDGKEYIKPEKL 1822
DB 278 GAYTKILEVMTKKGKLGTFYRVAFFYGO-----FPEDDGKEYIKPEKL 323
QY 1823 TPLSEISQRLKLYSDKGSENVKMIODSGVKNPKDLSKXAYVIOVTHVIFPFPEKELQ 1882
DB 324 TGLSEISRLKLYLXGKKGSTENVKMIODSGVKNKEDPKAHIOVTHVYKVPFDKELTE 383
QY 1883 RKTPEERSHNIRRFEMEPFTQTGKRGQGVBECKRRTILTAHCFFPYKGRIPMYQOH 1942
DB 384 RKTPEERSHNISRFFEAPYTLSGKQGCIEECKRRTILTSNSFPYKGRIPINCEQ 443
QY 1943 TDLPPIEVALIDEMSKVAELRQLCSAEVDMIKLOLKQGSVSVOVNAQPLAARAPLDD 2002
DB 444 INLKPIDVATEIDKTAELQKSCSTDVMDIQLQKQGSVSVOVNAQPLAARAPLDD 503
QY 2003 TNYKRPDNKTKLLEKVFQVEACGQALANVERLIXEDQLEOYEBEMKANREMAKELSE 2062
DB 504 SQASKYPPKXVSELKDMRKRTIQACSLALEMBELIKEDQYEHYEGKLSNRMVKEISD 563
QY 2063 IMHEO--ICPLEKTSVLP---NSLHIFNAISGT 2091
DB 564 IHEQASIRILOEDTMSPMMSNTLHVCAISGT 597

RESULT 10
Q921Y6 PRELIMINARY; PRT; 417 AA.
ID Q921Y6;
AC Q921Y6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ20220.
GN AA959601.
OS Mus musculus (Mouse).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC009134; AA009134.1; -  
 SO SEQUENCE 417 AA; 48431 MW; 6FCCEC3A90E059678 CRC64;

Query Match 19.1%; Score 2094; DB 11; Length 417;  
 Best Local Similarity 94.4%; Pred. No. 1,9e-125;  
 Matches 407; Conservative 6; Mismatches 4; Indels 14; Gaps 1;

QY 1677 VFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMELQCADGLMKARVELIND 1736  
 DB 1 MFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMELQCADGLMKARVELIND 60  
 QY 1737 IYKLIPIYKRRDFERLAHLYDTLHRAVSKYEVHSGRLLGTFFRVAFGQAAYOC 1796  
 DB 61 IYKLIPIYKRRDFERLAHLYDTLHRAVSKYEVHSGRLLGTFFRVAFGQ----- 114  
 QY 1797 TBSETDVEGFEEDGKEYIYKPKLTPLSEISORLLKLYSKFGSENNVMIDSGKNP 1856  
 DB 115 -----GFEEDEGKEYIYKPKLTPLSEISORLLKLYSKFGSENNVMIDSGKNP 166  
 QY 1857 KOLDSKAYIQVHVHPFPEKELOERKTEFERSHNIIRPFMPPTONGKRGQVEBOC 1916  
 DB 167 KOLDSKAYIQVHVHPFPEKELOERKTEFERSHNIIRPFMPPTONGKRGQVEBOC 226  
 QY 1917 KRRTIIITAHCFPYVKRIPVYQHTDLPNIEVAIDEMSKYAEELROLCSSAEVDMIKL 1976  
 DB 227 KRRTIIITAHCFPYVKRIPVYQHTDLPNIEVAIDEMSKYAEELROLCSSAEVDMIKL 286  
 QY 1977 QLLQSSVSQVNAAGPLAYARAFLDNTNTRYPDNRYKLYKEFRQPEACGALAVNER 2036  
 DB 287 QLLQSSVSQVNAAGPLAYARAFLDNTNTRYPDNRYKLYKEFRQPEACGALAVNER 346  
 QY 2037 LIKEDOLEYOEMKANYREMAKELSEIMHOICPLEKTSVLPNSLHIFNALSCTPTSTM 2096  
 DB 347 LIKEDOLEYOEMKANYREMAKELSDIMBOICPLEKTSVLPNSLHIFNALSCTPTSTV 406  
 QY 2097 VHGMTSSSVY 2107  
 DB 407 VQGLTSSSVY 417

RESULT 11  
 ID Q9GM45 PRELIMINARY; PRT; 425 AA.

AC Q9GM45;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE Hypothetical 48.6 kDa protein (Fragment)  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN PARIETAL LOBE;  
 RA Oseada N., Hida M., Kuenda J., Tanuma R., Iseki K., Hirai M., Teroo K.,  
 RA Suzuki Y., Sugeno S., Hashimoto K.;  
 RT Isolation of full-length cDNA clones from macaque brain cDNA  
 RT libraries.  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB049841; BAB16727.1; -  
 KW Hypothetical protein  
 FT NON\_TER 425 425  
 SO SEQUENCE 425 AA; 48574 MW; E05B25BEC9984003 CRC64;

Query Match 19.1%; Score 2085.5; DB 6; Length 425;  
 Best Local Similarity 93.8%; Pred. No. 6.7e-125;  
 Matches 408; Conservative 2; Mismatches 2; Indels 23; Gaps 1;

QY 1358 MHAARIQQLGSLDNLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDLSFTLAFKN 1417  
 DB 1 MHAARIQQLGSLDNLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDLSFTLAFKN 60  
 QY 1418 QLLADHGNHPLMKKYPDVYLCFLOKHQSEFTLAKVFTLRSLIYKPSFVEGRADMCQA 1477  
 DB 61 QLLADHGNHPLMKKYPDVYLCFLOKHQSEFTLAKVFTLRPLIYKPSFVEGRADMCQA 120  
 QY 1478 LCYEILKCNLSLSSIRTEASQLLYFLMRNNDYTGKSPYTHQLVIISQLIADYVG 1537  
 DB 121 LCYEILKCNLSLSSIRTEASQLLYFLMRNNDYTGKSPYTHQLVIISQLIADYVG 180  
 QY 1538 ICGTRFOOSLTIINNCANSRLIKTSPSSDYKDLTKIRIYVLMATQMKHEHNDPMLV 1597  
 DB 181 ICGTRFOOSLTIINNCANSRLIKTSPSSDYKDLTKIRIYVLMATQMKHEHNDPMLV 240  
 QY 1598 DLQYSLASVASTPELRKRTWDSMARIHVKNQDLSAAMCYVHTALVAEYLTREAVQM 1657  
 DB 241 DLQYSLASVASTPELRKRTWDSMARIHVKNQDLSAAMCYVHTALVAEYLTREAVQM 295  
 QY 1658 BPPLPHSHSACLRSGVFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMEL 1717  
 DB 296 -----GMFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMEL 337  
 QY 1718 LEQCADGLMKARVELIADYKLIPIYKRRDFERLAHLYDTLHRAVSKYEVHSGRR 1777  
 DB 338 LEQCADGLMKARVELIADYKLIPIYKRRDFERLAHLYDTLHRAVSKYEVHSGRR 397  
 QY 1778 LIGTFRVAFGQA 1792  
 DB 398 LIGTFRVAFGQVS 412

RESULT 12  
 ID Q9BMX9 PRELIMINARY; PRT; 696 AA.

AC Q9BMX9;  
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)  
 DE B125M24.1 (K1AA1058 protein) (Fragment).  
 GN B125M24.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wilson S.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL391237; CAC36036.1; -  
 FT NON\_TER 1 1  
 FT NON\_TER 696 696  
 SO SEQUENCE 696 AA; 81097 MW; ASBCBF055A94F7A8 CRC64;

Query Match 18.4%; Score 2007; DB 4; Length 696;  
 Best Local Similarity 53.0%; Pred. No. 1.5e-119;  
 Matches 390; Conservative 122; Mismatches 164; Indels 60; Gaps 10;

QY 600 VNSSTIPKQPE-TCSTKPTTFEVEBEFPCIPKXHTOPTITTNHLYVPKLYKIDSQVSF 658  
 DB 2 ITTSVPLKPEKQCN--ITVEVEBEFPEMTKCYPTIYKHLVYVPLQKIDSQKTF 59  
 QY 659 AKARNIAICIFKDSDEDSQPLKCIYRPGGVPFTRSAFAVYHHNPNFPEYDIKTEL 718  
 DB 60 AKARNIAICVFERDSDESDALKCIYKPGSFTTNVAVVSHHNPPEYDIKTEL 119  
 QY 719 PTQLHEKTHLLITFEHVSQDSSKSTKRDVETQVGSWLPPLKDGRTVTSQHIPLVS 778

[illegible]

RESULT	13
Q8WX11	
ID	PRELIMINARY; PRT; 534 AA.
AC	Q8WX11;
DT	01-MAR-2002 (TREMblrel. 20, Created)
DT	01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT	01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE	BH12804.1.2 (Novel protein, isoform 2) (Fragment).
GN	BH12804.1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Lawlor S.;
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL391280; CND21443.1; ..
FT	NON TER 1 1
FT	NON TER 534 534
ST	SEQUENCE 534 AA; 61426 MW; 8A0FD929A5ADD41F CRC64;
Query Match	18.2%; Score 1985.5; DB 4; Length 534;
Best Local Similarity	63.3%; Pred. No. 2.4e-118;
Matches 383; Conservative	76; Mismatches 71; Indels 75; Gaps 2;
QY	1463 PPSFTEERADWCALCYEIIKCCNSKUSITFEASQLLFLMRNNPDTGKSKFVTRTHL 1522
DB	1 FPSAPFKRNVNCAFCYEVLLKCCCTSKSI STNEASALLYLMRRNNFEYTKRTEFLRTHL 60

```

OY 1523 QVITISVOLADVVDIGSTGRQOGLSTINNCANDRLIXHTSPSSDYDILTKIRITLMA 1582
Db 61 QITIAVSQOLADVALSGSRFOESLPIINNFPANSDRPMKATAPFAEAVDILTKIRITLMA 120
OY 1583 TAQKHEHNDPEMLVDIOYSLAKSYASTPELAKTLWDSMARIHVONGDLSAAACVYHVT 1642
Db 121 TAQKHEHNDPEMLIDIOYSLAKSYASTPELAKTLWDSMAKHVONGDPSAAACVYHVA 180
OY 1643 ALVAEYILTRKEAVQWEPPLPHSHSACLRRSGGVFROGCTAFRVITPNIDEBASNMEDV 1702
Db 181 ALVAEPLHRKK-----LFPNGCSAFKKITPNIDEBGAMKEDA 217
OY 1703 GMDVHNEDVLMELLFQCADGLMKARVELADIYCLIIPIYKRRDPERLALHYOTLH 1763
Db 218 GMDVHVHSEVELLELLEQCVGDLMKARVELISELIVPIYKRRSF----- 266
OY 1763 RAYSKVTEVMHSGRLLIGTVRAVAFQGAQVQPTDSETVDEGFPEDBDGKEYIYKEPKL 1822
Db 267 -----ESFPEBDGKEYIYKEPKL 285
OY 1823 TPLSEISQRLIKLYSDKRGSENVKMIDSGKVNPKDLDSKVAYIIOVTHVIPFDEKELQE 1882
Db 286 TGLSEISRLVKLYGKRGTEVNVKIIDSDRVNAKEIDPKYAHIOVTVKRYPDOKELTE 345
OY 1883 RKTPESSHNRIRRMFEMFPYQTKRGQGVDEQCKRTIILALHCPRVYKRIIVMIOHH 1942
Db 346 RKTPESSHNRINSRVFAPYTLISCKKQGCIEOCCRRTIILTSSFPVYKRIIPINCEOQ 405
OY 1943 TDLNPIEVAIDEMSKYAELEOLCSSABVMDIKQLKLOQGSVSYOVNAGPLAVYARAFLLD 2002
Db 406 INLKPIDVADTEIDOKTAELDKLCCSSIDVDMIQLOKLQGSVSYOVNAGPLAVYARAFLLD 465
OY 2003 TINTKRYPDNKKVLLKEVYFQVFEACGQALAVNEBLIKEDOLEYOEMKANRENAKELSE 2062
Db 466 SQASKRYPKKVSSEIKOMFRKFIQCSIALBELNERLIKEDQVEYHEGKSNFRDVKELSD 525
OY 2063 IMHEQ 2067
Db 526 IMHEQ 530

```

```

RESULT 14
ID 045377 PRELIMINARY; PRT; 2374 AA.
AC 045377; 045399;
DC 01-JUN-1998 (TRMBLrel. 06, Created)
DT 01-NOV-1998 (TRMBLrel. 08, Last sequence update)
DR 01-JUN-2001 (TRMBLrel. 17, Last annotation update)
DE F22G12.5 protein.
GN F22G12.5
OS Caenorhabditis elegans.
OC Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida; Ehabditioidea;
OC Rhabditidae; Pelodderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard N.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnescoigh R., Anderson K., Baynes C., Berke N.,
RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Duthin R., Faveilo A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Luthmurtz J., Lloyd C., Mcmurtry A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ricken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock U., Wilkinson-Sproat V., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

[illegible]

Db	797	SLTNLIYTPALAKYDSOKAFSKARNISCTVRFYRG--EZAIPKAMVDNRSAAGP-XCI	853
Oy	696	SAPAAVHHHONPEFYDEIKIELEPTOLHEKHLLITFEFVSCNDSKSGSTKRKDVETOV	755
Db	854	SSFCVAQHQQONPVFGEEMKAQLPLNLTSTDSHLLFFSFHSISVAGNS--NLVASESTETPI	911
Oy	756	GYSWLEPL--KDGVMVNSECH--IPUSANPSGUYLGVOELGWMR--HYGEIIRVMGGK	808
Db	912	GYSWLEPLWKCDRLVMEWDEBFPLPAIADLPNYIRSKPTMGKGDSETTSEVMWD-QK	970
Oy	809	PLKISTVLSTVYVYTOQHNLNFEFOCYOKTESGQALAGNELVKYLSIHAHEGVMTAFL	868
Db	971	PLFVRILRIVSVETTTAKLOTFFQACDRL--SAKGIIIGDAADSLKLLSDVPFRLLLYL	1028
Oy	869	PTIINOLFVLTTRATQOEVAVNTRVIIHVVAQCHESGLSHLSYKYAVK---AEPY	924
Db	1029	PVIGRFLPALPOAPTEQOLATATLRISISIDMCCONGKPOVVRPFVSHPSDSVAEQF	1088
Oy	925	VASEKYVHEELTSMSTTLIKPSADFTSN-----KLKYSWFEFDVLIKMAOHLIENS	979
Db	1089	VSHR-TTIVSAICHLPTLMKE---LQSEEQALYLQULVLTDAVIKMAQTMCSQ	1143
Oy	980	KVLLRNQRFAPASYAHAVEYVVMNLPHITQKFRDNPBASKNANSLAVTKRCFTMDR	1039
Db	1144	LNTKSQRFRFOEILQMGQVLEVGVPQIVLKHMEBESRCAMIALAYFRFAMSFVDR	1203
Oy	1040	GFVFKOINNTYISGAPDDBPKTLEFYKKEPLRVYCNHNYIPLNIPM-----	1085
Db	1204	GVERWIFHYISRLDDTDYRALRDYKTDLLEILCLHEHNHVLNPLVLIINAAQIQRLNYS	1263
Oy	1086	-----PFGKG-----RIQ	1093
Db	1264	GGVVDYQMGQTNGTGLSGCYCSLRSRSTVVVNIIPASGSGFLSRFPNQIFNPTLETNETD	1323
Oy	1094	RYOQLODYSITDEFCKNNHVLGILLAEVGALOFEFVRLIAISVLKNLLIKSPDDR	1153
Db	1324	RYASCSGWMLSPFAPQNHPIVGLIMBELVACIRETDYRRKRPISLRNLAKHSPDRY	1383
Oy	1154	ASRSHQARIATLYPLFGLILENTQYINVRVS-----EPVNAQMTVXDESLAL	1203
Db	1384	GDMITQRIAMLYAPVLTFALPDHHERDSEIIDVDVATPIGYRSFPISSGGLP-----	1436
Oy	1204	PANVPLVTPQKSTLUNSLNKHLDLGAISGASPYTSTPNINSVRNDSRGLSTD--	1260
Db	1437	PSSSTTTLTSLKRWIVNR--YTDVQOQFSG--SPIRTSPP-VPSTQTPSPRPQMPPPA	1491
Oy	1261	SGNLSLPRNSKXSJDKHQOOSTLGNVVRCDLDOSEIKSLMCFYLILKXSDALF	1320
Db	1492	SGKTIPTPPPS-----SSTPLVEKLTBEIDODILICCVFIQNRPKILA	1535
Oy	1321	TYW--NKASTSE-----LMDF--TISEVCLHQF-----QYMKRYIAR	1355
Db	1536	ALMTENDQANAEMKIRLLELIVDFRLEMALQNVGIYEYQKIMLYLKKRYRGKEHALR	1595
Oy	1356	TGMHAKLOOL-----GSLNLSL---FNHSYGHSDMD-----VHQSLLE	1393
Db	1596	RTAANSKTRSLFTLINLBSRISSGSTPNRISAEEMMAAGABEDDPSGCTNPIPRVLOL	1655
Oy	1394	ANITETVCLPDLDTLSLFTLAFKNQOLAD--HGHP-----LMKRVFVUYLCELOKHOSET	1447
Db	1656	VNISQEVALLVDVAQ---TFMAQLASSQRRHRCPSQDALFSLMLALHLRIADENHSET	1711
Oy	1448	AKNOVFTALNSLYKEPSTYE-GRADMCALCYEILIKCNCKSLISRTESQOLYFLMR	1506
Db	1712	VRLHVIAGLALFVNLFFARLFEGGPRLEPLMYLIEKVLIQWASRLPAVOAAMALLQILR	1711
Oy	1507	NNFP-----YNGKSFVARTHLOVIVISQOLADVIGIGTR-----FQO	1545
Db	1772	NGIEVAQGTFRASOVLAAQSVSBSSKXIQOATRKGVSSERLGRGSGQCVALLAQFO	1830
Oy	1546	SLGSIINNCANSDRLI-----KHTSPSSVDYDQTLTKIRTVLMAVTAQMKENDPEM	1595



```
Qy 1642 TALVAEYLTRKAVOWEPPLPHSHSACLRSRGVFRQCTAFRVYTPNIDEASMMED 1701
Db 1108 AALVAEYLSMLB-----DHSY-----LPVGSVGFQNISSVLEESVYSED 1147
Qy 1702 VGMQD-----VHFNEDVLMELLEQCADGLMKAREYELIADYKLIPIYEKRDPERL 1754
Db 1148 TUSPDEBGVCAGQYFTESGLVGLLEQAAELFSTGLEYETVNEVYKLVIPLEAHRERFKL 1207
Qy 1755 AHLVDTLHRAVSKVTEVMSGRRLGTYPRAFFGQAQYQFTDSEVDVEGFEDDGKE 1814
Db 1208 TLTHSKLQRAFDSTVUNKDH--KRMFGTYFRVGFSGK-----FGDLDEQE 1250
Qy 1815 YIYKEPKLTPLSSEISQRLKLYSDKFSSENVKMIQDSGKVNPKDLSKYAVIQTVHVPF 1874
Db 1251 FVYKEPATITKLPEISHLEAFYGCFCGAEFEVIXDSTPVDKTKLDPNKAYIQITFVEPY 1310
Qy 1875 FDEKELOERTEPERSHNIRRFMEFPTOTGKQGVBEQCKRITILTALHCFPYVKR 1934
Db 1311 FDEYEMKDRVYFEKJFNLRPFMYTTPFTLEGRPRGELHEOYRRNTVLITMHAFPYIKTR 1370
Qy 1935 IPVWQHHTDNLPIEVAIDEMSKVAELRQLCSSAEVDMIKLOKLQGSVSVQVNAEPLA 1994
Db 1371 ISYIQKEBEFVLPIEVAIEDMKKKTTLQAVAINQEPDAMLOMVLQGSVGATVNOGPLE 1430
Qy 1995 YARAFPLD--TNTKRY-PDNKVKLLKEVFRQFEACGQALAVNERLIKEDOLEYQSEMKR 2051
Db 1431 VAQVFLAEIPADPKLYRHHNKRL---CFKEFIMRCGEAVEKNKRLITADQREYQOELKC 1487
Qy 2052 NYREMAKELSEIMHEOI 2068
Db 1488 NYNKLKENLRPMIERKI 1504
```

Search completed: July 14, 2003, 18:21:10  
Job time : 106.749 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 / Search time 26.0723 Seconds  
(without alignment)  
3843.329 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898  
Sequence: 1 MRLGAAVAALGRGRAPAS.....INKYLEATGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_101002.\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3898	100.0	752	AAE11888	Angiogenesis assoc
2	3840.5	98.5	751	AAE11888	Human GTP-binding
3	2521.5	94.7	729	ABE60766	Drosophila melanog
4	2275.5	58.4	442	AAV01167	Polypeptide fragme
5	2113.5	54.2	409	ABE89506	Human polypeptide
6	1489	38.2	692	AAU35975	Helicobacter pylori
7	1454	37.3	700	AAU35467	Haemophilus influe
8	1431	36.7	741	AAU40036	Propionibacterium
9	1427	36.6	704	AAE15891	E. coli proliferat
10	1427	36.6	704	AAU34770	E. coli cellular p

11	1412	36.2	705	22	AAE90301
12	1406	36.1	702	22	AAU36265
13	1387	35.6	695	23	ABE48403
14	1374	35.2	704	22	AAU38372
15	1371.5	35.2	692	23	ABP30759
16	1352.5	34.7	691	22	AAU33481
17	1352.5	34.7	693	19	AAU80722
18	1352.5	34.7	693	22	AAU48888
19	1352.5	34.7	693	22	AAU37640
20	1352.5	34.7	693	22	AAU01093
21	1351.5	34.7	715	22	AAU36850
22	1351.5	34.7	715	22	AAU37143
23	1346.5	34.5	692	23	ABP27487
24	1338	34.3	693	22	AAU33742
25	1318	33.8	705	23	ABP40511
26	1315	33.7	702	20	AAU37794
27	1308	33.6	709	23	ABE56515
28	1178.5	30.2	651	20	AAU35171
29	1175	30.1	245	22	AAU20463
30	1107	28.4	223	21	AAE42032
31	944	24.2	191	21	AAE42257
32	896	23.0	207	22	AAU20620
33	784.5	20.1	1093	22	ABE62821
34	772	19.8	719	22	AAU38875
35	692.5	17.8	735	22	AAE96105
36	686	17.6	332	18	AAE20914
37	568.5	14.6	641	13	AAE30089
38	541.5	13.9	842	21	AAV43637
39	541.5	13.9	842	21	AAV43630
40	539	13.8	639	23	ABP30306
41	539	13.8	644	23	ABP27486
42	538.5	13.8	514	23	ABP28786
43	535.5	13.7	514	23	ABP30078
44	530	13.6	619	9	AAE80147
45	529.5	13.6	523	23	ABE53656

## ALIGNMENTS

RESULT 1	
AAE11888	
ID	AAE11888 standard; Protein; 752 AA.
XX	
AC	AAE11888;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Angiogenesis associated hmt-elongation factor G (hEF-G) protein.
XX	
KW	Angiogenesis associated protein; AAF; cytosolic; cardiac; gene therapy; ophthalmological; vulnary; myocardial infarction; macular degeneration; diabetic retinopathy; angiogenesis; wound healing; prophylactic vaccine; rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation; cancer; therapeutic; diagnostic; hmt-elongation factor G; hEF-G.
OS	Unidentified.
XX	
PN	WO200170808-A2.
XX	
PD	27-SEP-2001.
XX	
PF	22-MAR-2001; 2001WO-US09609.
XX	
PR	22-MAR-2000; 2000US-191134P.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Rastelli LK, Gerritsen M;
XX	
DR	WPI; 2001-602775/68.
XX	
DR	N-PSDB; AAD19117.

XX Novel angiogenesis associated polypeptides and polynucleotides encoding  
PT the polypeptides, useful for modulating angiogenesis and for treating  
PT tumors and cancers -  
XX  
PS Claim 1; Page 15-18; 159pp; English.  
XX  
CC The invention relates to angiogenesis associated proteins (AAP) and their  
CC corresponding cDNA molecules, which are useful for modulating  
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting  
CC wound healing, for example after organ transplantation, and in the  
CC treatment of tumours, myocardial infarction, cancers, diabetic  
CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
CC AAP proteins and DNA's are useful in potential prophylactic and  
CC therapeutic applications implicated in a variety of disorders including  
CC those related to angiogenesis, and also in diagnostic applications.  
CC AAP cDNA is also useful in gene therapy. The invention also relates to  
CC a method for screening a tissue sample for tumourigenic potential. AAP  
CC proteins are used to screen drugs or compounds that modulate AAP activity  
CC or expression as well as treating disorders characterised by insufficient  
CC or excessive production of AAP or production of AAP forms that have  
CC decreased or aberrant activity compared to the wild type protein, or  
CC modulate biological function that involve AAP. The present sequence  
CC is hmt-elongation factor G (hmf-G) protein which is an angiogenesis  
CC associated protein (AAP) of the invention. Hmt-EF-G is strongly induced  
CC in an in vitro model of angiogenesis and is also used as an attractive  
CC therapeutic target to treat disease states associated with mitochondrial  
CC dysfunction.  
XX  
SQ Sequence 752 AA;  
Query Match 100.0%; Score 3898; DB 22; Length 752;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLGAAVAALGRGAPASLGWQRKQVNWKACRWSSGGVTPNEKIRNIGISAHDSGKT 60  
DB 1 MRLGAAVAALGRGAPASLGWQRKQVNWKACRWSSGGVTPNEKIRNIGISAHDSGKT 60  
QY 61 TLTERVLYYTGRIAKHNEVGKGVGAVNDMSLEORGTITTSATFTMKQVNNITD 120  
DB 61 TLTERVLYYTGRIAKHNEVGKGVGAVNDMSLEORGTITTSATFTMKQVNNITD 120  
QY 121 TPGHVDFTIEVERALRYLDGAVVLCVAGVGQCOTVTNNRQMRVYVPELTFTNKIDRMG 180  
DB 121 TPGHVDFTIEVERALRYLDGAVVLCVAGVGQCOTVTNNRQMRVYVPELTFTNKIDRMG 180  
QY 122 TPGHVDFTIEVERALRYLDGAVVLCVAGVGQCOTVTNNRQMRVYVPELTFTNKIDRMG 180  
DB 122 TPGHVDFTIEVERALRYLDGAVVLCVAGVGQCOTVTNNRQMRVYVPELTFTNKIDRMG 180  
QY 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERAIYFGDPSQIVRYGIRPA 240  
DB 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERAIYFGDPSQIVRYGIRPA 240  
QY 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERAIYFGDPSQIVRYGIRPA 240  
DB 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERAIYFGDPSQIVRYGIRPA 240  
QY 241 ELBAATDHRQELIECVANSDEQLGEMFLEEKIPTSIDKLAIIRBATLKRSFPPVLSGS 300  
DB 241 ELBAATDHRQELIECVANSDEQLGEMFLEEKIPTSIDKLAIIRBATLKRSFPPVLSGS 300  
QY 241 ELBAATDHRQELIECVANSDEQLGEMFLEEKIPTSIDKLAIIRBATLKRSFPPVLSGS 300  
DB 241 ELBAATDHRQELIECVANSDEQLGEMFLEEKIPTSIDKLAIIRBATLKRSFPPVLSGS 300  
QY 301 LKNKGVOPLDAVLEVLNPFSEVONTAIIINKDSDSEKTKILNNSRHNHPVGLAFPL 360  
DB 301 LKNKGVOPLDAVLEVLNPFSEVONTAIIINKDSDSEKTKILNNSRHNHPVGLAFPL 360  
QY 301 LKNKGVOPLDAVLEVLNPFSEVONTAIIINKDSDSEKTKILNNSRHNHPVGLAFPL 360  
DB 301 LKNKGVOPLDAVLEVLNPFSEVONTAIIINKDSDSEKTKILNNSRHNHPVGLAFPL 360  
QY 361 EVRFQGLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDADMEASTEEVYADICA 420  
DB 361 EVRFQGLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDADMEASTEEVYADICA 420  
QY 361 EVRFQGLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDADMEASTEEVYADICA 420  
DB 361 EVRFQGLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDADMEASTEEVYADICA 420  
QY 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSKNDLEKPSKIGFTEDDP 480  
DB 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSKNDLEKPSKIGFTEDDP 480  
QY 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSKNDLEKPSKIGFTEDDP 480  
DB 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSKNDLEKPSKIGFTEDDP 480  
QY 481 FKVPFTENKETVYISGGEHLHLEIYAQRLEREYGCCTGPKVARETITAVPDPFTH 540  
DB 481 FKVPFTENKETVYISGGEHLHLEIYAQRLEREYGCCTGPKVARETITAVPDPFTH 540  
QY 481 FKVPFTENKETVYISGGEHLHLEIYAQRLEREYGCCTGPKVARETITAVPDPFTH 540  
DB 481 FKVPFTENKETVYISGGEHLHLEIYAQRLEREYGCCTGPKVARETITAVPDPFTH 540  
QY 541 KKGSGAGQYGVKIVGLEPLDPEDYTKLEFSDTFPSGNI PKQFPAVERKGLDACCKGPL 600  
DB 541 KKGSGAGQYGVKIVGLEPLDPEDYTKLEFSDTFPSGNI PKQFPAVERKGLDACCKGPL 600

DB 541 KKGSGAGQYGVKIVGLEPLDPEDYTKLEFSDTFPSGNI PKQFPAVERKGLDACCKGPL 600  
QY 601 SGHKLSGARFVYQDDAHMWDSNETSPTRAGBALKOLANATLILPIMAVEVAPNE 660  
DB 601 SGHKLSGARFVYQDDAHMWDSNETSPTRAGBALKOLANATLILPIMAVEVAPNE 660  
QY 661 FQGVYAGINRRHGYITGQDGVEDFTLYADVPPLNDMGYSTELNSCTEGKGYTMEYSR 720  
DB 661 FQGVYAGINRRHGYITGQDGVEDFTLYADVPPLNDMGYSTELNSCTEGKGYTMEYSR 720  
QY 721 YQPCLPSTOEDVYINKYLEATGQLPVKKGAKN 752  
DB 721 YQPCLPSTOEDVYINKYLEATGQLPVKKGAKN 752  
RESULT 2  
AAB68560  
ID AAB68560 standard; Protein; 751 AA.  
AC AAB68560;  
XX  
DT 19-APR-2001 (first entry)  
XX  
DE Human GTP-binding associated protein #60.  
XX  
DE Human; guanosine triphosphate binding associated protein; GTP; GBAP;  
XX inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;  
XX autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;  
XX osteoporosis; psoriasis.  
XX  
OS Homo sapiens.  
XX  
PN WO200105970-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 19-JUL-2000; 2000WO-US19698.  
XX  
PR 19-JUL-1999; 99US-0144595.  
XX 23-AUG-1999; 99US-0150460.  
XX 15-OCT-1999; 99US-0159849.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Yue H, Tang YT, Bandman O, Hillman JL, Lai P, Au-Young J;  
PI Reddy R, Yang J, Baughn MR, Lu DM, Azimzal Y, Patterson C;  
XX WPI; 2001-091972/10.  
DR N-PSDB; AAF58360.  
XX  
PT New guanosine triphosphate-binding associated protein (GBAP) and their  
PT encoding nucleic acids, useful for treating and/or diagnosing diseases  
PT associated with GBAP expression, such as cancer, diabetes and asthma -  
XX  
XX Claim 1; Pages 178-180; 233pp; English.  
XX  
CC The present invention relates to novel human guanosine triphosphate  
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their  
CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences  
CC of the present invention are useful for treating a variety of disorders  
CC including inflammation, AIDS, Addison's disease, anaemia,  
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,  
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and  
CC psoriasis.  
XX  
SQ Sequence 751 AA;  
Query Match 98.5%; Score 3840.5; DB 22; Length 751;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 742; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 MRLGAAVAALGRGAPASLGWQRKQVNWKACRWSSGGVTPNEKIRNIGISAHDSGKT 60  
DB 1 MRLGAAVAALGRGAPASLGWQRKQVNWKACRWSSGGVTPNEKIRNIGISAHDSGKT 60





QY 644 LCTLEPIMAVEVAPNEFOGVIAGINRRHGVTGGDVEDYFTLYADVPNDMFYSTE 703  
 DB 621 WQLEBPIMLVEVAPNEFOGAVMGHLSKRGIITGTEGEGFTVVAEVLNDMPFGYAGE 680  
 QY 704 LRSECTGKGEYTMESRYOCPLESTOEDVINKYLEATGQLPVKKGAKN 752  
 DB 681 LRSTGQKGEFTMEYSRYSPLPDVODQIVROYESQGLAQDPKKKKKN 729

RESULT 4  
 ID AAY01167 standard; Protein; 442 AA.  
 AC AAY01167;  
 XX  
 DT 18-MAY-1999 (first entry)  
 XX  
 DE Polypeptide fragment encoded by gene 10.  
 XX  
 KW Human; secreted protein; gene therapy; protein therapy; tissue; cancer;  
 KW tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;  
 KW developmental abnormality; foetal deficiency; Alzheimer's disease;  
 KW cognitive disorder; schizophrenia; immunological disorder; mood disorder;  
 KW immune deficiency disease; respiratory disorder; arthritis; skeletal;  
 KW haematopoietic disorder; neural; osteoporosis; metabolic disorders;  
 KW cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9901020-A2.  
 PD 14-JAN-1999.  
 XX  
 PF 30-JUN-1998; 98MO-US13608.  
 XX  
 PR 12-SEP-1997; 97US-0058663.  
 PR 01-JUL-1997; 97US-0051381.  
 PR 01-JUL-1997; 97US-0051480.  
 PR 12-SEP-1997; 97US-0058598.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Carter KC, Endress GA, Peng P, Rosen CA, Ruben SM;  
 DR N-PSDB; AAX22120.  
 XX  
 XX MPI; 1999-105683/09.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, immune deficiency diseases or blood  
 PT disorders  
 XX  
 PS Disclosure; Page 17; 179pp; English.

The invention relates to nucleic acid sequences (AAX22111 to AAX22134)  
 encoding human secreted proteins (AAY01135 to AAY01158). The secreted  
 protein gene sequences are deposited with the ATCC under deposit number  
 ATCC 209118. Host cells comprising recombinant vectors containing the  
 nucleic acid sequences are used for the recombinant production of the  
 secreted proteins. The polynucleotide and amino acid sequences are useful  
 for are useful for preventing, treating or ameliorating medical  
 conditions e.g. by protein or gene therapy. Pathological conditions can  
 be also diagnosed by determining the amount of the new polypeptides in a  
 sample or by determining the presence of mutations in the new  
 polynucleotides. Specific uses are described for each of the  
 polynucleotides, based on which tissues they are most highly expressed  
 in, and include developing products for the diagnosis or treatment of  
 cancer, tumour, developmental abnormalities and foetal deficiencies,  
 autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,  
 schizophrenia, immunological disorders, immune deficiency diseases  
 (AIDS), mood disorders, respiratory disorders, arthritis, asthma,  
 haematopoietic disorders, neural disorders, skeletal disorders,  
 osteoporosis, metabolic disorders, cardiovascular disorders, endocrine

CC disorders or gastrointestinal disorders. The polypeptides are also useful  
 CC for identifying their binding partners. The present sequence represents a  
 CC polypeptide fragment encoded by a gene of the invention (see descriptor  
 CC line for gene number).  
 XX  
 SQ Sequence 442 AA;  
 Query Match 58.4%; Score 2275.5; DB 20; Length 442;  
 Best Local Similarity 98.6%; Pred. No. 5.4e-186;  
 Matches 437; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 310 LDAVLEIYPNPEVONVAILNKKDSSKEKTKLNMSSRNHSPFGLAFPLEVGRFGQLT 369  
 DB 1 LDAVLEIYPNPEVONVAILNKKDSSKEKTKLNMSSRNHSPFGLAFPLEVGRFGQLT 60  
 QY 370 YVRSYQGLKKGDDTYNTRTRKVRLORLARHADMMEASTREYVAGDICALFGIDCASG 429  
 DB 61 YVRSYQGLKKGDDTYNTRTRKVRLORLARHADMMEASTREYVAGDICALFGIDCASG 119  
 QY 430 DFTDKANSGLSMESIHVPDPVSIAMKPSNKNLDEKSKIGRTREDPTPKVYFDTEN 489  
 DB 120 DFTDKANSGLSMESIHVPDPVSIAMKPSNKNLDEKSKIGRTREDPTPKVYFDTEN 179  
 QY 490 KETVISGMBELHLEYAQRLEHREYCCPTGKPKAFRETTIAPVPFPDTHKQSGGAGQ 549  
 DB 180 KETVISGMBELHLEYAQRLEHREYCCPTGKPKAFRETTIAPVPFPDTHKQSGGAGQ 239  
 QY 550 YGKVIQVLEPDPEDYTKLEFSDERTFGSNIPQPFAPAVEKGLDACEGPLSGHLSGLR 609  
 DB 240 YGKVIQVLEPDPEDYTKLEFSDERTFGSNIPQPFAPAVEKGLDACEGPLSGHLSGLR 299  
 QY 610 FVLQDGAHMDVDSNEISFIRAGEGALKQALANATLCILEPIMAVEVAPNEFOGVIAGI 669  
 DB 300 FVLQDGAHMDVDSNEISFIRAGEGALKQALANATLCILEPIMAVEVAPNEFOGVIAGI 359  
 QY 670 NRRAGVITGGDVEDYFTLYADVPNDMFYSTELRSCTEGGEYTMESRYOCPCLPSTQ 729  
 DB 360 NRRAGVITGGDVEDYFTLYADVPNDMFYSTELRSCTEGGEYTMESRYOCPCLPSTQ 419  
 QY 730 EDVINKYLEATGQLPVKKGAKN 752  
 DB 420 EDVINKYLEATGQLPVKKGAKN 442

RESULT 5  
 ID ABB89506 standard; Protein; 409 AA.  
 AC ABB89506;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 1882.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnervary; anticonvulsant; anticholinergic; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;

XX WPI; 2002-122018/16.  
DR N-PSDB; ABL89915.  
XX  
XX  
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX  
XX  
PS Class 11; SEQ ID NO 1082; 2081pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
CC (ABL89404-ABL90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 409 AA:  
Query Match 54.2%; Score 2113.5; DB 23; Length 409;  
Best Local Similarity 98.8%; Pred. No. 3.6e-172;  
Matches 405; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 343 NMSRRNSHPVGLAFPLEVGRFGQLTYVRSYQGLKKGITTYNTRKKVRLQRLARNH 402  
DB 1 NMSRRNSHPVGLAFPLEVGRFGQLTYVRSYQGLKKGITTYNTRKKVRLQRLARNH 60  
QY 403 ADMMASTBEVYAGDICALFGIDCAGDTFTDKANGLSMESIHVDPIYISIAMKSNKN 462  
DB 61 ADMMA-DVEEVYAGDICALFGIDCAGDTFTDKANGLSMESIHVDPIYISIAMKSNKN 119  
QY 463 DLEKFSKIGRFTREDPTFKVYFDTNKKETVISMGELEHLEIYAOBLEEYGGPCITGKP 522  
DB 120 DLEKFSKIGRFTREDPTFKVYFDTNKKETVISMGELEHLEIYAOBLEEYGGPCITGKP 179  
QY 523 KVAFRBTITAPVPDFTHKKQSGAGQYGVIGLEPLDPEDTYKLEFSDFTFGSNIPKQ 582  
DB 180 KVAFRBTITAPVPDFTHKKQSGAGQYGVIGLEPLDPEDTYKLEFSDFTFGSNIPKQ 239  
QY 583 FVPVAVKGLDACEKPLSGHLSGRFLYLDGSAHMDVDSNLSFIRAEAGALKQALANA 642  
DB 240 FVPVAVKGLDACEKPLSGHLSGRFLYLDGSAHMDVDSNLSFIRAEAGALKQALANA 299  
QY 643 TICIIEPIAAVEVVAENEFQGOVIAGINRRHGVITGODGVEDFTLYAADVPLINDMGYST 702  
DB 300 TICIIEPIAAVEVVAENEFQGOVIAGINRRHGVITGODGVEDFTLYAADVPLINDMGYST 359  
QY 703 ELRSCTEGKGEMTEYSRYQPCLPSTQEDVINKYLEATQQLPVKKGAKAN 752  
DB 360 ELRSCTEGKGEMTEYSRYQPCLPSTQEDVINKYLEATQQLPVKKGAKAN 409  
RESULT 6  
ID AUU35975 standard; Protein; 692 AA.  
XX AUU35975;  
AC  
XX  
DT 14-FEB-2002 (first entry)  
XX

DE Helicobacter pylori cellular proliferation protein #288.  
XX  
XX Antisense; prokaryotic cellular proliferation protein;  
XX antibiotic; antibacterial; drug design.  
XX  
OS Helicobacter pylori.  
XX  
PN WO200170955-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-20727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
DR N-PSDB; AAS53834.  
XX  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX  
PS Example 3; Seq ID No 11568; 511pp; English.  
XX  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 692 AA:  
Query Match 38.2%; Score 1489; DB 22; Length 692;  
Best Local Similarity 45.7%; Pred. No. 2.3e-116;  
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;  
QY 42 PNEKIRNIGISAHIDSGKTTLTERVLYYTGRIARKEVKGKDGAVAMDSELERGIGT 101  
DB 6 PLNRRIRNGIAHAIDAGKTTTSERLFTYGVSHKIGEV--HDG-AAATMDWMEQEKRGIT 62  
QY 102 IOSAATFTMKQDVNINIITDTSRHVDFTIEVERALAVDGAVALVCAVGVCQCTTVARQ 161  
DB 63 ITSAAITFCWKDQHQINLIDTPEHVDFTIEVERSMKVLGAASVFCVGVQVQPSFTVMQ 122  
QY 162 MKRVNVPFLTFINKLDRGMSNPARALQOMRSKLNNTAFMOIPMGLBGNFKGIVLISER 221  
DB 123 ANKIVGPRIVFVNRKDRIGANFNVENQIKORLKNPVPINIPJABDTFTGVIDLVQMK 182  
QY 222 AIYFDGDFSQIVRYG--EIPALRAAATDHRQELIECVANSDBQJENFLEKIPISIDL 279

```

Db      183 AIWMNNE-TWGAQYDVEIIPSDLEKAKQYREKLEVAVAEQDBALMEKYLGGELDIERI 241
Qy      280 KLAIRRAATLKRSFTPLGSALEKNGVPLDVALELNPSPVQVAYAILANKDDSEKRT 339
Db      242 KGIKTCGLNMSVPMGLCSSFNKGVQTLDAVIDLPAPELVNI----KSIDRTEB 297
Qy      340 KILMNSRRNHSHPVGLAFLPVEGRF-GQLTVYRSYOGELKKDITVTRTRKVVRL 398
Db      298 EHVYKSS--DDGEFALAFKIMDPVQGLTFRVRKGLSSSYNNSTKKKEKVGRL 355
Qy      399 ARNHADMEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDVISIANK 457
Db      356 LKMSHMKRE-DIKEVYAGEICAFVGLKDTLTGDTLDCDERKA-VLERMSEPEPEVHIAVE 413
Qy      458 PSKNDLEKSKKIGFTREDPFKYYFDTENKETVYISGMEGLHLEIYQRLEREKGC 517
Db      414 PRTKAOEKKKVALGKLAEDPSFRWTOEQTGLIGGELHLEIYDRLRERKVEA 473
Qy      518 ITGKPYAFRETTTAPVPDFTHKKQSGAGQYKAYIGVLEPLDPEDYTLKESDETFGS 577
Db      474 EIGQPQVAFRETTIRSSVSKENKAKQSGRGQYGHVFIKLEPEKPG--SGYEFVNISSG 531
Qy      578 NIPKQFVPAVEKGFLLDACEKGPLSGHKLGLRFLVLDGAHHMVDNSIEISFIRAGCALQ 637
Db      532 VIPKEYIIPAVDKGIOEAMONGVLAGYPAVDKVTLYDGSYHVDSSSEMAFKIAGSMAPKE 591
Qy      638 ALANATLCTLEPIMAVEVVAAPNEFOGOVYAGINRRHGVTIGQGVADYFTLVADVPALM 697
Db      592 ASRAAPVPLEPMPKKEVEVEPEYMGVDVIGDLRRRGQINSMDRLGLKIIVNAPVLEVM 651
Qy      698 FGYSLELRSCTECKGEYTMESRYOPLPSTOEDVINK 735
Db      652 FGYSTDLRSATQGRGYSMEFHDYGVPSIAKEIYEK 689

RESULT 7
AAU35467 standard; Protein; 700 AA.
ID      AAU35467;
XX      AAU35467;
XX      14-FEB-2002 (first entry)
DT      14-FEB-2002 (first entry)
XX      Haemophilus influenzae cellular proliferation protein #108.
DE      Haemophilus influenzae cellular proliferation protein;
XX      Antisense; prokaryotic cellular proliferation protein;
XX      antidiotic; antibacterial; drug design.
XX      Haemophilus influenzae.
OS      Haemophilus influenzae.
XX      MO200170955-A2.
PN      MO200170955-A2.
PD      27-SEP-2001.
XX      27-SEP-2001.
XX      21-MAR-2001; 2001WO-US09180.
PF      21-MAR-2001; 2000US-191078P.
XX      21-MAR-2001; 2000US-191078P.
PR      23-MAY-2000; 2000US-206848P.
XX      23-MAY-2000; 2000US-206848P.
PR      26-MAY-2000; 2000US-207727P.
XX      26-MAY-2000; 2000US-207727P.
PR      23-OCT-2000; 2000US-242578P.
XX      23-OCT-2000; 2000US-242578P.
PR      27-NOV-2000; 2000US-253625P.
XX      27-NOV-2000; 2000US-253625P.
PR      22-DEC-2000; 2000US-257931P.
XX      22-DEC-2000; 2000US-257931P.
PR      16-FEB-2001; 2001US-269308P.
XX      16-FEB-2001; 2001US-269308P.
XX      (ELIT-) ELITRA PHARM INC.
PA      (ELIT-) ELITRA PHARM INC.
PI      Haeselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
XX      Yamamoto RT, Xu HH;
XX      WPI; 2001-611495/70.
DR      N-PSDB; AAS53326.
XX      N-PSDB; AAS53326.
XX

```

```

PT      New polynucleotides for the identification and development of
XX      antibiotics, comprise sequences of antisense nucleic acids -
XX      Example 3; Seq ID No 11060; 511pp; English.
Cc      The invention relates to antisense inhibitors of genes essential to
Cc      prokaryotic cellular proliferation, their use in identifying the
Cc      genes, their use in the discovery of novel antibiotics, the essential
Cc      genes themselves and the encoded proteins. The prokaryotes used are
Cc      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
Cc      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
Cc      invention is also useful for the identification of potential new targets
Cc      for antibiotic development. The antisense nucleic acids can also be used
Cc      to identify proteins used in proliferation, to express these proteins,
Cc      and to obtain antibodies capable of binding to the expressed protein.
Cc      The proteins can be used to screen compounds in rational drug discovery
Cc      programmes. The antisense nucleic acid sequence is also useful to screen
Cc      for homologous nucleic acids which are required for cell proliferation in
Cc      a wide variety of organisms. The present sequence represents an
Cc      essential prokaryotic cellular proliferation protein.
Cc      Note: The sequence data for this patent did not form part
Cc      of the printed specification, but was obtained in electronic
Cc      format directly from WIPO at
Cc      ftp.wipo.int/pub/published_pct_sequences.
XX      SQ
XX      Sequence 700 AA:
XX      Query Match 37.3%; Score 1454; DB 22; Length 700;
XX      Best Local Similarity 45.1%; Pred. No. 2.3e-115;
XX      Matches 318; Conservative 124; Mismatches 235; Indels 28; Gaps 13;

Qy      42 PNEIRNIGISAHIDSGKTTLTERVLYYTGRIAKNHEVYKQGVAVNDSELERQGIT 101
Db      6 PIERRNIGISAHIDAGKTTTERLFTYGVSHKIGEV--HGC-AAITDMWEGDERGIT 62
Qy      102 IQSAATFTWKKDVN-----INIITDGHVDFTIEBERALVLDGAVLVLCAGGVQCC 154
Db      63 ITSAAITFAWMSGMSQOPPOHRIINVIDTGHVDFIVEERSMKVLGAVWVYCAVGVQVQ 122
Qy      155 TMTVNRQMKRYVPLFTFLINKLDRMGSNPARALQOMRSKLTNNNTAFMOIPMGLGNPKFI 214
Db      123 SETVMRQANKYEVPRIPAFVKNKDRGTGANFLRVBOLKTRLGNAIPLQVPAEENFTCV 182
Qy      215 VDLIEERAI-YFDGFSQIVRYGEIPABLRRAATRHROELBVCVANSDEQLCEMLEBEKI 273
Db      183 VDLIMKALINWENAOQMTFTFEYVPAWMOQDCEMRQVLVEAALASELMEKYLGCSD 242
Qy      274 PSISDLKLAIRPATLKRSFTPLGSALEKNGVPLDVALELNPSPVQVAYAILANKDD 333
Db      243 LTBEEIKSALRQVLANEILVTCGSAFKNKGVQAMLDVAVAYLPAPTDIP--AIKGINP 300
Qy      334 DSKEKTKILMNSRRNS--HPFVGLAFLPVEGRF-GQLTVYRSYOGELKKDITVTRNR 390
Db      301 DETE-----GERASDBEPSPSLAFKAIADPFVGNLTFPRVYGVINSGDTVLNSVRQ 353
Qy      391 KKVRLQRLAAMHADMEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPD 449
Db      354 KKEKRGRIYQMANRRE-EIKEVRAAGDIAAALGLDVTGDTLC-AIDAPILMEMEPFE 411
Qy      450 PVISIAMKPSKNDLEKSKKIGFTREDPFKYYFDTENKETVYISGMEGLHLEIYQRL 509
Db      412 PVISVAVEPKTADDEKKGALGRLAOEDPSFRVHTDESGETIISGMEGLHLDIVIRM 471
Qy      510 EREYGCPTITGKPYAFRETTIAPV-PRDFTHKKQSGAGQYKAYIGVLEPLDPEDYTKL 568
Db      472 KKEFVEANIGKPYVRETTIRTVNDVVGKIAKQSGGQYGHVVIDLYPLDPBG-PGY 530
Qy      569 ERSDETFGSNIPKQFVPAVEKGFLLDACEKGPLSGHKLGLRFLVLDGAHHMVDNSIEIS 628
Db      531 EFWNEIKGVIPGEYIIPAVDKGIOELKSGPLAGYPAVDKVTLYDGSYHVDSSSELANK 590
Qy      629 RAGEGALQALANATLCTLEPIMAVEVVAAPNEFOGOVYAGINRRHGVTIGQGVADYFTL 688

```

Query Match	Score	DB	Length
Best Local Similarity	44.1%;	Pred. No. 2.3e-113;	
Query Match	36.7%;	Score 1431;	DB 22;
Best Local Similarity	44.1%;	Pred. No. 2.3e-113;	Length 741;

Matches	308; Conservative	137; Mismatches	232; Indels	22; Gaps	13
QY	45	KIRNIGISAHIDSGKTTTTERVLYTTGRIIAKQHEVYKDGCVGVAMVDSMELEFORGITTS	104		
Db	52	KVRNIGIMAHIDAGKTTTTERILLYTTGKSRYKIGEV--HDG--AAATMDWKEOEORGITTS	108		
QY	105	AAATMMKDVANINIIDPGRVDFTEVEERARLVLDGAVLVLCAVGGVQCGMTTVAROMGR	164		
Db	109	AAATTFMHMDTOINIITDPGHVDFPTEVERSRRLVDGAAVAFEDGAVAGVEQSMTWKQAK	168		
QY	165	YNVDFELFINKLDRMGSPAPALAOOMRSKLTNNHTAFMOWIPMGLEGNPKGIVLIEERAIY	224		
Db	169	YGVFRICITINLDRGASFPHCVKTIKRLNAIPVLQDLPICAEQOFMGIYDLVENNACT	228		
QY	225	FDGDFSGQIVRY--GEIPALRLAATDNRBOELIECVANSDEQLGEMFLEE-KIPISIDLKL	281		
Db	229	WRGETELCAHYETEDIPADMKDEAPAAABAEMLETAENDEBFMEYLSDPDVAVTVDQVKA	288		
QY	282	AIRRAATLKRGSTTPVFLGSAIKKKGVQPLLDVLETLPRPSEVQNTAIVLANKDDSKETKI	341		
Db	289	AIRRGVLSAATAVTAACGTISFNKKGVQPLLDIAVDYLPSPDTPAISGFRKGDESEV---	344		
QY	342	LMNSSRHSHFVGGLAPFLLEVQ--RRGOLTYVRSYGGELKKGGTIVNTRTRKKVRLORLAR	400		
Db	345	-LERRPCDDEPLSTIAFKIASDPRHKLTLFRAVYSGVLHAAGDQVLNATYKKKERIGKITQ	403		
QY	401	MHADMEASTEEVYVAGDICALFGI-DCASGDTFTDKANSGLSMESIHVPDPVISTAMKPS	459		
Db	404	MNAKRE-EIISIGGMCICAMVGLKDTTGTETLSDQNTP--IILSEMTTPAPVIEQALBPK	461		
QY	460	NKNDLEKFSKIGIGRTREDPTFFKQVFDTENKETVISINGBELHLIYQRLERBYGCPCT	519		
Db	462	SKADQEKISNAIORLVEEDPTFRVATDEBTQCTIVAGNGELHLDAFIDRMKEEPFHEANI	521		
QY	520	GKPVAPFETITAPV-PEDFTNKKKSGGAGQGVKIVGLVLEPRDPTKLEPSDETFGSN	578		
Db	522	GKPOVAHEITLRKPEKEEYTHKKQGTGSGPARVIAIIEPKEP--FGYEIVNNAVTSGR	579		
QY	579	IPKQFVPAVEKGPLDACEKGPLSGHKLSGLRFVLQDGAHHMVDNSNBSIFIRAGEGLAKQA	638		
Db	580	IPKXIYIPSVDAVGQESMGFGVLAVGPVEDVAVKTLLDGAVHEVDSSEBNAPKLAGSNAFKXA	639		
QY	639	LAANATLCLEPITMAVEVVAENVEFOGOVIATGINRRRHGVITGDQGVEDY--FTLYADVPLND	696		
Db	640	ARKANPGILBPLMAVEVTPPEDYLGTVIGDLSNRGQI--QEMVEHGNKVRAVLVPLSB	697		
QY	697	MEGYSTEIARCTBEKGGEYTMYSKRYQCLPSTQDEVIYVK	735		
Db	698	MFGYVGDLRSKTSQASYSMEFDSYGBPCSSVADEIIAK	736		
RESULT 9					
AAB15891					
ID	AAB15891 standard; Protein; 704 AA.				
XX	AAB15891;				
XX	05-OCT-2000 (first entry)				
DE	E. coli proliferation associated protein sequence SEQ ID NO:248.				
XX	Escherichia coli; E. coli; proliferation; inhibition; screening;				
KW	antimicrobial; bacterial growth; antisense therapy; antibacterial.				
OS	Escherichia coli.				
XX	MO200044906-A2.				
XX	03-AUG-2000.				
XX	27-JAN-2000; 2000WO-US02200.				
XX	27-JAN-1999; 99US-0117405.				
PR					

XX (ELIT-) ELITRA PHARM INC.  
 PA Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ,  
 PI Yamamoto RT, Xu HH;  
 DR WPI: 2000-514822/46.  
 DR N-PSDB; AAA65895.  
 CC Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 PS Claim 11; Page 178-180; 316pp; English.

XX Sequence 704 AA;

DB 534 YEFINDIKGVIPEYIPAVDKIGIQOLKAGELAGYPVVDGMGIRLHFGSHYDVSSELA 593  
 QY 628 IAGEGALQKOLANATLCTLEPIMAVEVAPNEFOGOVYAGINRRHGVITGQGVEDYPT 687  
 DB 594 KLAASIAFKEGFKKAKPVLLIEPIMKVEVETPEENTGVIGDSRRGMKKGSEVYTGK 653  
 QY 688 LVADVPPLNDMFGYSTELRSCTEGKGEYTWESYRPOCLPSTQEDVINKYLEATGQ 742  
 DB 654 HIAEVPLSEMFQVATQRLSTKGRASVTMEFLKYDEA-PS---NVAQAVIEARGK 704  
 RESULT 10  
 AAU34770  
 ID AAU34770 standard; Protein; 704 AA.  
 AC AAU34770;  
 DT 14-FEB-2002 (first entry)  
 DE E. coli cellular proliferation protein #351.  
 KW Antisense; prokaryotic cellular proliferation protein;  
 OS antibiotic; antibacterial; drug design.  
 OS Escherichia coli.  
 PN WO200170955-A2.  
 XX 27-SEP-2001.  
 PD 21-MAR-2001; 2001WO-US09180.  
 PF 21-MAR-2001; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,  
 PI Yamamoto RT, Xu HH;  
 DR WPI: 2001-611495/70.  
 DR N-PSDB; AAS52629.

Query Match 36.6%; Score 1427; DB 21; Length 704;  
 Best Local Similarity 44.2%; Pred. No. 4.7e-113;  
 Matches 316; Conservative 130; Mismatches 239; Indels 30; Gaps 15;  
 QY 42 PNEKINIGISAHIDGKTTLTERRVLYGRIKAKMEVKGKGVGVMSMEIERGRTT 101  
 DB 6 PIARYNIGISAHIDGKTTTERRILFYGVNKHIEV-HDG-AATMMMEQEOERGIT 62  
 QY 102 IOSAFTFTMWKQV-----NINIIDTPGHVDFTEVERALRYLDGAVLVCAVGVOO 154  
 DB 63 ITSAATTATKSGMAKQYEPHRIINIDTPGHVDFTEVERSMRVLDGAVMYCAVGVOPO 122  
 QY 155 TWTVNRQMKRYNVPFLTFINKLDRMGSPAPALQOMRSKLNHTAPMOJPMGLEGNPKGI 214  
 DB 123 SETVMQANKYKVPRIAFVNMKRMGNFLKVVNQIKTRIGANPVLQALIGMEHPTGV 182  
 QY 215 VDLIERBAL-YFPGDPSQIVRYGEIPEALRAAATDHRQELIECVANSDBOLGEMPLEBK 273  
 DB 183 VDLVKKKAIINMDADQGVTFEYEDIPADWVELNEMHONLIBSAAASEBELMEKYLGESE 242  
 QY 274 PSISDKLAIIRATLRSFPVFLGSLKNGVQPLDVLAVLEYLPSEPOVNAIILNKDO 333  
 DB 243 LTBELKAGLRQVNLNNEIILVTGSAFKNGVQAMLDVAIDLPSPVDP-AINGILD 300  
 QY 334 DSKEKTKIIMNSSRHNS--HPFVGLAPPLEVGRF-GQLTVRSYOGELKKGDTIYVTRR 390  
 DB 301 DGDQTE-----ARRHSDDEPFSALFKIATDPFVGNLTFRRYSGVANSQDVLNSVKA 355  
 QY 391 KKYRLORLARMAADMEASTEYVYADICALFEI-DCASGDTFTDYANSGLSMESHVHPD 449  
 DB 356 ARERFRIVOMHANKRE-EIKEVRADIDAAIQLKVTGTDTLCD-PDAPIILIERMEFPE 413  
 QY 450 PVISIMKPSNKNLDLKFSGKIGRFRBDPTFKVYEDTENKEKLVISGEGELHLEIYAOL 509  
 DB 414 PVISIVPEKTKADQEKGLALRLKALDEDSFVWMDSESNQITINGMGLHLDIIVDKN 473  
 QY 510 EREYGCPTIGKPKVAFRETTITAPV-PFPTTHKQSGAGQYKVGIVLEPLDPEDYTK- 567  
 DB 474 KKEFNVANVGKQVAVRETIRKQVTDVEGKAQSGRGQYHVIVDWPLEPGSNPBG 533  
 QY 568 LBSDBTFSGNIPKQVPAVEKGFLLDACEKPLSGHKLGLRFTVLDDGAHMYDSNEISF 627

QY 568 LBSDBTFSGNIPKQVPAVEKGFLLDACEKPLSGHKLGLRFTVLDDGAHMYDSNEISF 627

XX Example 3; Seq ID No 10363; 511pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 704 AA;  
 Query Match 36.6%; Score 1427; DB 22; Length 704;  
 Best Local Similarity 44.2%; Pred. No. 4,7e-113;  
 Matches 316; Conservative 130; Mismatches 239; Indels 30; Gaps 15;  
 QY 42 PNEKRNIGISAHIDSGKTTTTERVLYTGRIAKMEYKKGQGVGVAVNDMEIERGITT 101  
 DB 6 PIARYNNIGISAHIDAGKTTTTERILFYGVNHKIEV--HDG-AATMDMEQEGERGKIT 62  
 QY 102 IGSAAFTMMKOV-----NINIITDGHVDPTTEVERALRLDGAIVLCAVGVQOQ 154  
 DB 63 ITSAATTAFTWGSRAKQYEPHRINIIDTGHVDPTTEVERSRMRDLDAVWYCAVGVQOQ 122  
 QY 155 TITVNRQMKRYNVPFLTFINKLDRMGSPAPALQMRSKLNHTAEMQIPMGLEGFKFI 214  
 DB 123 SETWRQAKKQYKVPYLAFTVNRKMDRKANFLKYNVQIKTILSANPVQLAIGAEHFTV 182  
 QY 215 VDLIEBRAI-YPDGDSQIVRYGEIPAEIRAAATDRQELIECVANSDEQLGEMFLEEKI 273  
 DB 183 VDLVKKAIINMDADQGVTFEVEDIPADWVELNEMHQULIESAABASELMETKYLGE 242  
 QY 274 PSLSDKALIRATILKRSFTVPFLGSLAKKQYQPLLDVALEYLTPPSEVQNTAILNKD 333  
 DB 243 LIEABIKGALRQVLRVNNELIIVTCGSAFKNQVQAMLDAVIDLPSPIVDP-AINGILD 300  
 QY 334 DSEKTKIIMNSSRHNS--HPFVGLAFPLEVGRF-GQLTVRSYOGELKKGDTIYTRTR 390  
 DB 301 DKKDTP----AERHASDDEPFSALAFKIATDPFVNLIIFRYSGVANSQDVLVNSVKA 355  
 QY 391 KKVRLQRLARHADMMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPD 449  
 DB 356 ABERFRIYQMHANRKE-EIKEVAGDIAAIGLKQVTTGDTLCD-PDAPIILERMEPE 413  
 QY 450 PVIISAMKSNKNDLEKFSKIGRFRDPPTPKVYVDPTENKEVVISGMEHLIETAOQL 509  
 DB 414 PVISIAVEHKTADQKMGALGRALKEDPSFRVWDEBSNQIILAGMELHLIDIVDM 473  
 QY 510 EREYGCPTGKPKVAFRETITAPV-PFDTHKKSQSGAGQYGVKIVGLEPIIDPEDYTK- 567  
 DB 474 KBEFVNAVGKROVAYRETIRKQVTDVEGKAKQSGKQGVHVIDMPLFPGSNPKG 533  
 QY 568 LEFSDETFGSNI PKQFVPAVEKGFLDACEKPLSGHKLGLRFLVLDGAAHWNDSNEISF 627  
 DB 534 YEFINDIKGQVLPGEYIIPAVDKIQGLKAGPLAGVAVTWMGIRLHFGSYHDVDSLELAF 593  
 QY 628 IAAAGBALQALANATLCTLEPIMAVEVVAENFQGVLAIGINRRHGVTTGQDGYEDYFT 687  
 DB 594 KLAASIAFIEGFKKAPVLLPEIMKVEETPEENTGDVIGDLSRRRGMKGQSESVTVGK 653  
 QY 688 LVADVPVLMNPGYSTELRSGCTEGSKGYTMEYSYQPCLPSTQEDVINKYLEANTQ 742  
 DB 654 IHADEVPLSHMFGVATQLRSLITKGRASYMEFLKIDEA-PS--NVAQAVIEAR GK 704  
 RESULT 11  
 AAG90301  
 ID AAG90301 standard; Protein; 705 AA.  
 XX AAG90301;  
 AC 26-SEP-2001 (first entry)  
 DT  
 DE C glutamicum protein fragment SEQ ID NO: 4055.  
 XX  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX Corynebacterium glutamicum.  
 OS  
 XX EPI108790-A2.  
 PN  
 PD 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.  
 PF 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX (KYOW) KYOWA HAKKO KOGYO KK.  
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AA65520.  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 PS Claim 17; SEQ ID NO: 4055; 246bp + Sequence listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 705 AA;  
 Query Match 36.2%; Score 1412; DB 22; Length 705;  
 Best Local Similarity 42.7%; Pred. No. 9.1e-112;  
 Matches 301; Conservative 136; Mismatches 238; Indels 30; Gaps 12;  
 QY 45 KIRNIGISAHIDSGKTTTTERVLYTGRIAKMEYKKGQGVGVAVNDMEIERGITTQS 104  
 DB 7 KVRNIGIMAHIDAGKTTTTERILFYTGIRKNGE--HDG-GATDWMQGBERGKITTS 63  
 QY 105 AATFTMKQDVNNIITDGHVDPTTEVERALRLDGAIVLCAVGVQOQNTVNRQMR 164  
 DB 64 AAVTCFWMNNQVNIIDTGHVDPTTEVERSLRLDGAIVAFDQKGVPEQSRQVRQATK 123  
 QY 165 YVNPVLITINKLDRMGSPAPALQMRSKLNHTAEMQIPMGLENFQYVLDIEBRAIY 224  
 DB 124 YDVPRICTVNNKDKLGAFYFTVGTIEDRLKAPLWMLPFGAENFGVIDLLMKALT 183  
 QY 225 FPDGDSQIVRYG-----EIPAEIRAAATDRQELIECVANSDEQLGEMFLEEKIPIISD 278  
 DB 184 WRG-----VTPICTEATVEIRIPELADRAAEYREKLEFVAESDESLMEKYPGCELSIAE 239  
 QY 279 LKLAIRATILKRSFTVPFLGSLAKKQYQPLLDVALEYLTPPSEVQNTAILNKDQSKK 338  
 DB 240 IKAIRKRVNNSIIPYVCGTAYKNGKIQLDLDAVDELPSPLDGR---TKGTDVADP 295  
 QY 339 TKILNNSGRHNSHPFVGLAFPLEVGR-FGQLTVRSYOGELKKGDTIYTRTRKKVRLOR 397  
 DB 296 EKVLTTRKP-SDEEPLSALAFKIAHPFQKLPVLAISGKVPQGVANSTNKKKERIGK 354  
 QY 398 IARHADMMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPIYSIAM 456  
 DB 355 LFQMHAN-KENPVFAHNGNIYAFGLDKDTTGTGDTLCD-ANAPIILBMDPDPPIQYVAL 412  
 QY 457 KPSNNDLEKFSKIGRFRDPPTPKVYFDTENKTVISGMEHLIETAOQLREYVGP 516  
 DB 413 EPKTSDEKLGVAIQKLAEDPFTVHALDDESGQTVIGMGEHLIDVLVNRMKKEPFVY 472

QY	5171CTGKPKVAFREITTPV-PFDTHKKQSGAQYGVIGVLEPFLDE-----DVTKL 568
DB	473 ANIGDQVAYREITRRPVESLSTTHKKQYQSGQFAKVIITTPPYAPAEADLEEGESAIY 532
QY	559 EFSDETFGSNIPKQFPAVEKEFLDA CEKQPLSGHKLGLRFLYLDGAAHMDNSIEISFT 628
DB	533 KFNENATGVRPREYIPSDAGIQDMQVGFLLAGYPLVNVKATLEDDA YADVDSSEMAFK 592
QY	629 RAGEGALKQALNATLCTILEPIPAVAVVA PNEQGVYAGINRRHGYITQDQVEDFTL 688
DB	593 LMGSQAFKAAVAAKVLEPIPSVETITPEETMEGVIGVNSRRRQIASMDRAGAKLV 652
QY	689 YADVPIDMFQYSTBELRSCTECKGGEYTMESRRQPCLPSTQEDVI 733
DB	653 KAKVPLSCMFQYVGDRLSKTQGRANYSWPFDSIAEYPAVVAADVI 697
RESULT 12	
AAU36265	AAU36265 standard; Protein; 702 AA.
AC	AAU36265;
DT	14-FEB-2002 (first entry)
XX	Pseudomonas aeruginosa cellular proliferation protein #255.
DE	Pseudomonas aeruginosa cellular proliferation protein;
KW	antibiotic; antibacterial; drug design.
XX	Pseudomonas aeruginosa.
OS	WO200170955-A2.
PN	27-SEP-2001.
XX	21-MAR-2001; 2001WO-US09180.
PR	21-MAR-2000; 2000US-191078P.
PR	23-MAY-2000; 2000US-206848P.
PR	26-MAY-2000; 2000US-207727P.
PR	23-OCT-2000; 2000US-242578P.
PR	27-NOV-2000; 2000US-253625P.
PR	22-DEC-2000; 2000US-257931P.
PR	16-FEB-2001; 2001US-269308P.
XX	(ELITRA PHARM INC.
PA	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI	Yamamoto RT, Xu HH;
XX	WPI: 2001-611495/70.
DR	N-PSDB; MMS54124.
XX	New polynucleotides for the identification and development of
PT	antibiotics, comprise sequences of antisense nucleic acids -
PS	Example 3; Seq ID No 11858; 511pp; English.
XX	The invention relates to antisense inhibitors of genes essential to
CC	prokaryotic cellular proliferation, their use in identifying the
CC	genes, their use in the discovery of novel antibiotics, the essential
CC	genes themselves and the encoded proteins. The prokaryotes used are
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC	invention is also useful for the identification of potential new targets
CC	for antibiotic development. The antisense nucleic acids can also be used
CC	to identify proteins used in proliferation, to express these proteins,
CC	and to obtain antibodies capable of binding to the expressed proteins.
CC	The proteins can be used to screen compounds in rational drug discovery
CC	programmes. The antisense nucleic acid sequence is also useful to screen
CC	for homologous nucleic acids which are required for cell proliferation in

Query Match	36.1%	Score 1406	DB 22	Length 702
Best Local Similarity	43.6%	Pred. No. 2.9e-111		
Matches 308	Conservative 121	Mismatches 248	Indels 30	Gaps 13
<p>CC a wide variety of organisms. The present sequence represents an            CC essential prokaryotic cellular proliferation protein.            CC Note: The sequence data for this patent did not form part            CC of the printed specification, but was obtained in electronic            CC format directly from WIPO at            CC ftp.wipo.int/pub/published_pct_sequences.</p>				
Sequence	702 AA			
Qy	42	PNKIRNIGIAHIDSGKTLTTERLYTGTSLAKMHEKKGKGVANMDSMELEBORGIT	101	
Db	6	PIELRYNIGIAHVADAGKTTTTERLFTYGNHKGGEV--HDG-AATDMWVQEBORGIT	62	
Qy	102	IQSAATFTMKD-----VNINIIDPGHVDPIETEBARALVLDGAVLYLCAVGVOG	153	
Db	63	ITSAATTAFMGGSTKQFPHRYRNFNIDIPGHVDPIETEBARALVLDGAVVYSSGADGVEP	122	
Qy	154	QTMVTRNRMKRYNVPFLTEINKLDRMGNPAPALQOMRSKLNHTAFMQIEPGLGNFKG	213	
Db	123	QSEFTVWQANKXHYERLAVYNNKMRQGDPLFRVVAQIKQRLGHVVPQIALGISEENFSG	182	
Qy	214	IVDLIEERAIYF--DGDPSQIVRYGEIIPAELEAAADHNOELLECANDEQGEMLPEK	272	
Db	183	QIDIVKMKAIWNADQGSTYREBEIPELALAEWHAHNVAAEANDDELNKKYLEGE	242	
Qy	273	IPSIDLKLAIIRATYLKRSFTPVFGSLAKKKGVOPDLDAVLEYLPNPSEVONVAILNK	332	
Db	243	ELSLIEEIKAGIRQTLTANQIYPAVLGSSFFKKGVPLVDVAIDYIPARSEIE--AIRGD	300	
Qy	333	DSSEKTKIILNNSRH--NSHPFVGLAPFLEVGAF--GQLTYRSYQGLKKGDTTYNFT	389	
Db	301	PDDEK-----HDERHADDEPFSAIAFKIATDPVGTLTFAVYSGVLTSGDAVLNSVK	355	
Qy	390	RKRYRLQRLAMADMMEMASTEETVYAGTICALFGI--DCASGDTF--TDKANGLSMESIH	446	
Db	356	GKKERVGMVGMHANNQRD--EIKERVAGDIALIGMKDVTTGTLCAIDKP--IILERMD	411	
Qy	447	VPDVISIANKPSKNDLEKPSKGIGRFTREDPPKVFPTDENKETVISMGELEHLEYA	506	
Db	412	FPDVISIYVAVEPKTKADEKMGIALSKAQEDPSFRVKTDEDTAOTTISGMELHDIIV	471	
Qy	507	QRLREYVGCPTCKPKVAFRETTTAPVPFDTFNKKSQSGAQYQKVIQVLEPLDEDT	566	
Db	472	DRMRERFGEVANEICKQVAYARETIRNTCEISGKFRQSGGGRQGFHCWIRFAPAD--EGE	530	
Qy	567	KLEFSDETFGSNIPKQVPAVEKEGFLDACEGPLSGHKLGLGRFLPDGAAHMDVSNELS	626	
Db	531	GLEFHNVEVGVVIREFPIAOKGIEDQMONGVLAVGLIGLKATVYGSYHVDVDSSEVA	590	
Qy	627	FIRAGEGALKALNATLICLEPIEMAVEVVA--PNEQGVINGVINGNRHGVITQDQVEDVF	686	
Db	591	FKLIASMATKQLSKRGGAVALLEPMKKEVVPEDMGVMGMDLNNRRGLIQMEDTTPACK	650	
Qy	687	TLVADVPALDMFGSTELRSCTEGKEGYTMEYSRYQPLCPSTOEDVI 733		
Db	651	VIRAEVPLGEMFGYATIDVRSMSQGRASISMEFVRYAIVPASYABEIV 697		
<p>RESULT 13</p> <p>ID ABB48403 standard; Protein; 695 AA.</p> <p>XX ABB48403;</p> <p>XX AC</p> <p>XX 05-FEB-2002 (first entry)</p> <p>XX DE</p> <p>XX Listeria monocytogenes protein #1107.</p> <p>XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;</p>				







Db 65 SAATTAQWOGHARNVNIIDTGHVDFTEVORSLEVLDAVTVLDAQSGVEPQETETVWRQAT 124  
QY 164 RYVNPFLTFINKLDRMGSPARALQOMRSKLNNTA FMOIPMGLEGNFKGIVDLIEERAI 223  
Db 125 EYGVPRIVPANKNDKIGADFLYSVQSLHRLQANNAHP IOLPISSEDDFRGIIDLKCKAE 184  
QY 224 YPDGDFSQIVRYGELIPAELEAAATDHRQELIECVANSDEQLGEMFLEEKIP SIDSJLAI 283  
Db 185 IYNDLGTDLBEDIDPAEYVDQANEYREKULVEAVADTDEDLMMKYLEGEBETBELMAAI 244  
QY 284 RRATLRKSTFPVFLGSLANKKVQPLDAVLELYLPNPSEVQNTAII LKKODSKEKTKIIM 343  
Db 245 RKATINVEFPVLCGSAFKKGVQLMLDAVIDYLPSPLDI PAIKGINPTDDEE----- 298  
QY 344 NSGRHNSHPFVGLAFLVGRF-GOLTYVRSYOGELKKGDTIYNTRTRKKVRLQRLAH 402  
Db 299 TRPASDEBPALAFKIMTDPFVGRLPFRKVSGLVNSGSYVLTNSKGRERIGRILOMH 358  
QY 403 ADMMEASTEVEVYAGDICALFGI-DCASGDTFTDKANSGLMESIHVPDPVSIAMKPSNK 461  
Db 359 ANSRQ-EIETVYAGDIAAAGLKDITTGSLTDE-KSKVILBSIEVBPVIOLMVEPKSK 416  
QY 462 NDLEKFSKGI GRFTREDPTFKYFTDENKETVISGMEHLLEIYAO LBEREYCCPCTGK 521  
Db 417 AQQDKKGIALQKLABEDPTFRVETNVTGETVISGMGELHLDV LDRMKKEFFVEANVGA 476  
QY 522 PKVAFRETTAPVPFPDTHKKSGAGOVKIVGLEPLDBDYTKLEFSDETFGSNIPK 581  
Db 477 POUVSRETFRASTOARGFFRQSGGKQFGDV--WLEFTNERGKGREFENAI VGVVPR 534  
QY 582 QFVPAVEKGLDACERKPLSGHLSGLRFLVLODAHMYDSNEISFIRAGEGALKOALAN 641  
Db 535 EPIPAVEKGLVESMANGVLAGYPMVDYKAKLYDSYHDVDSSETAFKIASLALKEAKS 594  
QY 642 ATLCLIEPIMAVEVAPNEFQGOVIAGINRRHGVITGQDGVEDYFTLYADVPLINDMFGYS 701  
Db 595 ACPAILLEPMMLVTITAPEDNLAGDVMGHVTARGRVDGMEARGNTQVVRAFVPLAEMFGYA 654  
QY 702 TELRSGTEGKGTYMEYSRYQPLPSTQEDVINK 735  
Db 655 TVLRSATQCRGTMMVDPDHYEDVPKSVQEBIIRK 688

Search completed: July 14, 2003, 18:15:02  
Job time : 30.0723 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

# OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 3.44246 Seconds  
(without alignments)  
2620.804 Million cell updates/sec

Title: US-09-815-379-6  
Perfect score: 3898  
Sequence: 1 MRLGMAAVALGRGRAPAS.....INKYLEANGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1318	33.8	705	4 US-09-134-001C-5356	Sequence 5356, Ap
2	569.5	14.6	641	1 US-07-718-535-3	Sequence 3, Appl
3	569.5	14.6	641	1 US-08-161-999-3	Sequence 3, Appl
4	541.5	13.9	842	4 US-09-293-549-2	Sequence 2, Appl
5	541.5	13.9	842	4 US-09-293-549-4	Sequence 4, Appl
6	519.5	13.3	514	2 US-08-865-311-2	Sequence 2, Appl
7	519.5	13.3	514	4 US-09-315-720-2	Sequence 2, Appl
8	519.5	13.3	514	4 US-09-134-001C-3995	Sequence 3995, Ap
9	456.5	11.7	517	4 US-08-984-618-9	Sequence 9, Appl
10	445.5	11.4	517	4 US-08-984-618-7	Sequence 7, Appl
11	441	11.3	529	4 US-08-984-618-10	Sequence 10, Appl
12	440	11.3	529	4 US-08-984-618-8	Sequence 8, Appl
13	414	10.6	634	4 US-09-134-001C-3390	Sequence 3390, Ap
14	367	9.4	531	4 US-09-134-001C-3574	Sequence 3574, Ap
15	338.5	8.7	173	4 US-09-095-855-207	Sequence 207, App
16	338.5	8.7	173	4 US-09-205-426-207	Sequence 207, App
17	331.5	8.5	207	4 US-08-858-207A-345	Sequence 345, App
18	312	8.0	586	1 US-08-198-446B-19	Sequence 19, Appl
19	312	8.0	586	2 US-08-870-693-19	Sequence 19, Appl
20	217	5.6	408	4 US-09-140-466-4	Sequence 4, Appl
21	210.5	5.4	361	4 US-09-183-861-85	Sequence 85, Appl
22	210.5	5.4	361	4 US-09-022-765-85	Sequence 85, Appl
23	183	4.7	728	4 US-09-134-001C-4968	Sequence 4968, Ap
24	182.5	4.7	409	4 US-09-140-466-5	Sequence 5, Appl
25	178	4.6	409	4 US-09-140-466-3	Sequence 3, Appl
26	177.5	4.6	414	4 US-09-134-001C-5377	Sequence 5377, Ap
27	175.5	4.5	930	3 US-09-283-763-2	Sequence 2, Appl

28	175.5	4.5	930	4 US-09-574-912-2	Sequence 2, Appl
29	171.5	4.4	394	4 US-09-218-197-2	Sequence 2, Appl
30	168.5	4.3	462	1 US-08-299-351-1	Sequence 1, Appl
31	168.5	4.3	462	4 US-09-309-572-10	Sequence 10, Appl
32	168.5	4.3	462	6 5225348-1	Patent No. 5225348
33	160.5	4.1	462	2 US-08-371-377-18	Sequence 18, Appl
34	155	4.0	460	4 US-09-174-768-4	Sequence 4, Appl
35	154	4.0	394	4 US-09-140-466-6	Sequence 6, Appl
36	141	3.6	69	2 US-08-997-080-184	Sequence 184, App
37	141	3.6	69	2 US-08-997-362-184	Sequence 184, App
38	141	3.6	69	4 US-09-095-855-184	Sequence 184, App
39	141	3.6	69	4 US-09-324-542-184	Sequence 184, App
40	141	3.6	69	4 US-09-205-426-184	Sequence 184, App
41	137.5	3.5	410	4 US-09-140-466-2	Sequence 2, Appl
42	126.5	3.2	555	2 US-08-982-232-14	Sequence 14, Appl
43	125.5	3.2	2482	1 US-08-328-254-6	Sequence 6, Appl
44	125.5	3.2	3248	1 US-08-353-700-1	Sequence 1, Appl
45	125.5	3.2	3248	5 PCT-US95-16216-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-134-001C-5356  
; Sequence 5356, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR APPLICATION NUMBER: 1998-08-13  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5356  
; LENGTH: 705  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5356

Query Match 33.8%; Score 1318; DB 4; Length 705;  
Best Local Similarity 40.5%; Pred. No. 1.4e-121;  
Matches 284; Conservative 134; Mismatches 249; Indels 34; Gaps 9;  
QY 47 RNIGSAHIDSGKTLTRVLYYTGRIAKMEHVKGDGAVMDSMELEROGITIOSNA 106  
DB 23 RNIGMAHIDAKTTLTRILYTRIRIKIGFT--HEG-ASQMDMEGQDQGITITSNA 79  
QY 107 TETMKADVNIINIDPFGHVDFTIEVERALRVLDGAVLVCAVGVCOCOTVYNRQMKRYN 166  
DB 80 TTAQOQGRHVNIIDPFGHVDFTIEVERSLRVLDGAVTVLDQSGVGPOTETVWRATTYGV 139  
QY 167 VEFLEFINKLDMGNSNPALQOMSKLNHNFAQIOIMGEGNKGIYDLLEERAIYVD 226  
DB 140 VRIIVFVNMDLGNLFYSVSTLHRLQANAPQLPFGADDEFEALIIDLVEMKCFKTT 199  
QY 227 GDFSOIVRYGEIPALRAAATDHRQELIECVANSDEQGEMLFEKIPISIDLKLAIRRA 286  
DB 200 NDLGHEIDEIEIPDHKRAERARQQLLEAVAKENNDDLMKYLGPBEISVDELKQAIRQA 259  
QY 287 TIKRSFTYFGLSALKNGVOPLDVAIEYLPNBEVQ-----NYAILNKKDDSK 336  
DB 260 TTDVEFYFYLCTGTAFKNGVQMLNAVIDYFSPPLDVKPIICHRANNPDEEVAVKPDQSA 319  
QY 337 ETKTKLMSRNRNSHPFGLAEPLEVGPR-GQLTVVRSGEELKKGGDTIYNTRTKKVYL 395  
DB 320 E-----FAALAFKVMTPDVPYKLTFRVYSGTSSGYSVKNSSKKRREKV 364

QY 396 ORLARNHADMMEEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPEVSI 454  
Bb 365 GRLLQNHANSRQ-EIOTVYSGELIAAVGLKETGTOTLCEKND-IILSMEPPEVYIH 422  
QY 455 AMKPSNKNDLEKFSKIGRFTREDPTFKVYFTENKETVYISGGEHLLEYAORLEREG 514  
Db 423 SVBPESKADODKMTOLVKLOEEDPTFHATDETEQVLIIGMGELHLDLVDRMKERN 482  
QY 515 CPETKPKVAFREBITTAPVPFPTHKKGSGAGOGYKVLGVLEPIDPEYTLERSDEI 574  
Db 483 VECNVGAPMVSYNETFKOPQVOGKSRSGRGQYGDV--HLETPNETGGSFERNMI 540  
QY 575 FGSNPKOPFVPAVEKGFLDACEKGPLSGHRLSGLRVLQDGAHHMYDSNEISFIRAGEA 634  
Db 541 VGGVREVYIPVSEQGLKDMENGVLAVGLYLVKAKLFDGSHYDVDSSEMAFKIASLA 600  
QY 635 LKALANATLCTLEPTMAVEVAPNEFOGOVINGINRRHGVITGQGVEDYFLVAVPL 694  
Db 601 LKEAAKKCDPVILPEPMKVTIEMPEEYMGDIMGDVARRGRVDMGEMPRGNAQVYNAVPL 660  
QY 695 NDMFGYSTELRSTCEGKEYTMEYSRYQPLCPSTOEDVINK 735  
Db 661 SEMFGYATSLRSNTQGRGTYTTFDHYAEVPSIAEEIINK 701

## RESULT 2

US-07-718-535-3  
Sequence 3, Application US/07718535  
Patent No. 5322784  
GENERAL INFORMATION:  
APPLICANT: Salyers, Abigail A.,  
APPLICANT: Shoemaker, Nadja B.,  
APPLICANT: Nikolich, Mikeljon P.,  
TITLE OF INVENTION: Method and Materials For  
Introducing DNA Into Prevoecella ruminicola  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William Brinks Olds Hofer Gilson and Lione  
STREET: P.O. Box 10395  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM XT  
OPERATING SYSTEM: MS-DOS 3.31  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07718,535  
FILING DATE: 05-JUN-1991  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Marnell W.,  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3617/22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 321-4200  
TELEFAX: (312) 321-4239  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 641 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown

Query Match 14.6%; Score 569.5; DB 1; Length 641;

Best local Similarity 26.1%; Pred. No. 212e-47;

Matches 182; Conservative 120; Mismatches 287; Indels 109; Gaps 17;

QY 46 IRNIGISAHDSKTLTTERVLVYTGRIAMGHEVKKDGVAGVMSMELEROGITIGA 105

Db 3 IINGLIHAIDAKTSTVENLLFASGATEKCGCVNDG---TITDSMIEKRRGITVRAS 59  
QY 106 AFTFMKOVNINIIDTPGHVDFTIEVERALRYLDGAVLYLCAVGVOGOTMVRNOMKRY 165  
Db 60 TSIITMGVKNIIIDTPGHMDTIAEVERFKPKLDGAVLYLSKEDIQATKILFTTLQYL 119  
QY 166 NVPTFTINKLDRMGSPNAPALQOMRSKLNHTAFMOJPMGLEGNFKGIVDLIBERAIYF 225  
Db 120 QLPITLFIINKIDRAGVNLERLTLDKANLSGOVLFMQ-----NVVD-----GSYV- 164  
QY 226 DDDFQIYRYGERIPALPRAATDHRQELIECVANDEQLGEMFLEKIPSLSDLKLAIR 285  
Db 165 -----PV-----CSQTYIKEEYKFPVCHNDHDLERYLADSEISPDYNNYITIA 208  
QY 286 ATLKSPFPVPLGSLAKKKGVOPLDVALEYLNPSEYONVAILINKKDSKEKTKILMNS 345  
Db 209 LVAKAKVYFVLHGSAMFNIGINELDALITSLFPASVSNRL-----SGLYILIEHP 261  
QY 346 SRHNSHPVGLAFPLEVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLAMHADW 405  
Db 262 KGH-----KRSFLKIIDGSLRLDVDVRINDSEKFIKIKVLKTLINGQ- 302  
QY 406 MEASTEEVYAGDICALFGID-----CASGDTFTDKANSGLSMESIHPDPEVSI 453  
Db 303 REINDEVGANDIAIVEBDDDFRIGNYLAEPCLL-----OGLSHQ-----HPLK 348  
QY 454 IAMKPSNKNDLEKFSKIGRFTREDPTFKVYFTENKETVYISGGEHLLEYAORLEREG 513  
Db 349 SSVRPDRPERKSKVYLSALNTLMTIEDPSLSFINSISDELSLYLTOKEIIOQLIEERF 408  
QY 514 GCPCTTGKPKVARE-----TITAPVPDPTHKKGSGAGOGYKVLGV-LEPLDEPD 564  
Db 409 SVKHFDEIKTIYKRPVAKVKIKIIEVP-----PNPYWATIGLLEPLIGT 457  
QY 565 YTKLEFSDPTFGSNIPKQFVPAVEKGFLDACEKGPLSGHRLSGLRVLQDGAHHMYDSNE 624  
Db 458 GLQIR-SDISYG-YLNHSFONAVPEGIRMSCSG-LHGMETVDLKVTFQAEVYSPVSY 514  
QY 625 ISFIRAGGALKQALANATLCTLEPTMAVEVAPNEFOGOVAGINRRHGVITGQGVED 684  
Db 515 ADFRQUTPVYFRLAQOQGVLDLEMLYFELQIPQASSKAITDQKMSIEDISCNNE 574  
QY 685 YFTLVADVPLNDMFGYSTELRSTCEGKEYTMEYSRYQ 722  
Db 575 WCHIKGVPLNTSKOYASEVSYTGKLGIFMVKPCGYO 612

## RESULT 3

US-08-161-999-3  
Sequence 3, Application US/08161999  
Patent No. 5674733  
GENERAL INFORMATION:  
APPLICANT: Salyers, Abigail A.,  
APPLICANT: Shoemaker, Nadja B.,  
APPLICANT: Nikolich, Mikeljon P.,  
TITLE OF INVENTION: Method and Materials For  
Introducing DNA Into Prevoecella ruminicola  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William Brinks Olds Hofer Gilson and Lione  
STREET: P.O. Box 10395  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM XT  
OPERATING SYSTEM: MS-DOS 3.31  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/161,999  
FILING DATE: 02-DEC-1993

```
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/718,535
/ FILING DATE: 05-JUN-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Crook, Manneli M.
/ REGISTRATION NUMBER: 31,071
/ REFERENCE/DOCKET NUMBER: 3617/22
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (312) 321-4200
/ TELEFAX: (312) 321-4299
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 641 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: unknown
/
US-08-161-999-3
```

Query Match 14.6%; Score 569.5; DB 1; Length 641;  
Best Local Similarity 26.1%; Pred. No. 2,2e-47;  
Matches 182; Conservative 120; Mismatches 287; Indels 109; Gaps 17;

```
QY 46 INNIGISAHIDSGKTLTERVLYYTGRIAMHEVKGKGVGAVMDSMELEROGITIOGA 105
DB 3 INNLGILAHIDAGKTSVTEHLFASGATEKCGCVNDG---TITDSMDIEKRGITVIRAS 59
QY 106 APTFMKQDNNINIDPBGHVDTIEVERALRVLDGAVVLCANVGVOCCMTVNRQMKRY 165
DB 60 TSIITWNGVKCNIDTPGHMDPTAEVERFKMDGAVLLISAEGLQAOQTKLLENTLQ 119
QY 166 NVPLFTPKLDRMGSNPARALQOMRSKLNHTAFMQIPMGLEGNFKGIYDLIEERATP 225
DB 120 QPTTIFINKIRBAGNLERLYDIKANISQDVLFWQ-----NNYD---GSYV--- 164
QY 226 DGDFOIYVGEIPAEPLRAATDHRQELIECVANSSEQLSEMFLEKIPISIDLKAIIR 285
DB 165 -----PV---CSQTYIKSEYEFVCHNDNIEERYLADEISPADYNTIILA 208
QY 286 ATLKSFTEPLGSAKKNKGVOPLDVLEYPSEVQNVYALINKKDSKEKTKILMNS 345
DB 209 LYAKAKVYVLHGSAFNIGINELDAITSPILPPASVSRL-----SSYLYKLEHDP 261
QY 346 SHNSHPVGLAFPLEVGRFGQLTYYRSYQGELEKKGDTIYNTTRKKVRLQRLARHADM 405
DB 262 KGH-----KRSPLKIIDGLRLADVAINSEKFIKIKULKITINOG- 302
QY 406 MEASTEEVYAGDICALFGID-----CASGDTFTDKANSGLSMESIHPDPVIS 453
DB 303 REINDEVGANDIAIVEDMDDFRIGNVLAEPCLL-----QGLSHQ-----HPALK 348
QY 454 IAMPKKNKNDLEKFSGIGRFTREDPTFKYYPTEKKEVYISGMGELHLEIYARLEREY 513
DB 349 SSVRPRPERSKVISALNTLWIEDSPSINSYDELEISLYGLOKEIITQTLLEERF 408
QY 514 GCPCTIGKPKVAFRE-----TITAPVPDFTHKQSGAGQYKVIQV-LEPLDPD 564
DB 409 SVKVNDELKITYKEKRPVKKNKIQLQEVF-----PNPYATIGLTLEPLPLGT 457
QY 565 YTKLEFSDTEPGSNIKQFVPAVEKGFLLDACEKGPLSGHKLGLRFLVDGAAHMYDSNE 624
DB 458 GLQIE-SUISYG-YLNHSFQNAVFEGIRMSQSG-LHGWEVTDLKVTFQAEYSPVSYP 514
QY 625 ISPIRAGEKALQALNATLCTLEPIMAVEVAPNPFQCYAVAGINRRRGVITGQGVGD 684
DB 515 ADPRQLTPVYFRALAQOQSGVDLLEPMLYELQIPQAASEKAITDLOKMSSEIEDISCNNE 574
QY 685 YFTLVADVPNDMFPGYSTELRSCTEGKEGYTMEYSRYQ 722
DB 575 WCHIKKQVPLNTSKOYASBVSYYTKGLGIFMWKPCGYQ 612
```

RESULT 4

```
US-09-293-549-2
/ Sequence 2, Application US/09293549
/ Patent No. 6440409
/ GENERAL INFORMATION:
/ APPLICANT: G. Todd Milne
/ APPLICANT: Gerald Fink
/ TITLE OF INVENTION: A METHOD FOR EFFICIENT AND HIGHLY
/ TITLE OF INVENTION: SELECTIVE CONTROL OF MICROORGANISMS
/ FILE REFERENCE: 50078/008002
/ CURRENT APPLICATION NUMBER: US/09/293,549
/ CURRENT FILING DATE: 1999-04-16
/ EARLIER APPLICATION NUMBER: 60/082,089
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 842
/ TYPE: PRT
/ ORGANISM: Saccharomyces cerevisiae
/
US-09-293-549-2
```

Query Match 13.9%; Score 541.5; DB 4; Length 842;  
Best Local Similarity 23.7%; Pred. No. 2,2e-44;  
Matches 201; Conservative 133; Mismatches 278; Indels 235; Gaps 30;

```
QY 46 INNIGISAHIDSGKTLTERVLYYTGRIAMHEVKGKGVGAVMDSMELEROGITIOGA 105
DB 19 VAMSVIAHVDHSGKTLTDSLVRAGIIS-----AAKGAEARFTTRKDEGRGITTIST 73
QY 106 APTFMKQDNN-----INIDPBGHVDTIEVERALRVLDGAVVLCANV 148
DB 74 A-ISLYSEMSDDEVKIKQKTDGNSFLINLIDSFGHVFSSEVTAALRVTDGALVVVDPI 132
QY 149 GGVQOQMTVNRQMKRYVNPFLITFKLDRMGSNPARALQOMRSKLNHTAFMQIPMGLE 208
DB 133 EGVCAQTEVTLKQALGERIKRPVVVINKVDR-----ALLELQVSKE 172
QY 209 GNPFGKIVDLIEERAI-----YFGDPSQIVRYGEIPAE-----LRAAATDH 249
DB 173 DLXQFARTVESVNVIVSTYADEVLDQVY---PARGVAFSGSLHGMAFTRQPAITRY 229
QY 250 ROEL-IECVANSDEQLGEMFLEKIPISIDLKAIIRATLKSGFT-----PVF-LGSAIK 302
DB 230 AKKFGVDKAKMMDRLMGDSFFNPRTKWTNKOTDAEGKPLERAFFNFIIDPIFLFTALM 289
QY 303 N-----KGVQ-----PLLDAYLR---YLPPSPVQ 324
DB 290 NPKKDEIPVLEKLEIVLGGEKDEKALLKVVNRKFLPADALLEMIIVLHLPSPVTAQ 349
QY 325 NY-----ATLNKKDQSKETKILMNSRRNSHPFVGLAFPLEVGRFGQLTYY 370
DB 350 AYRAQOLYEGPADDANCIALKNCDPKADLMLYSKNVFIS-----DKGRF---YAF 397
QY 371 VASYGELKKGDTIYNTTRKKVR-----LOQLARHADMMEASTEE 412
DB 398 GAVFPGTYVSG-----QKVRIGPNVYVPGKKDPLFIKAIQGVLMGMGRFVE-PIID 447
QY 413 VYAGDICALFGID---CASGDTFTDKANSGLSMESIHPDPVISIAMPKSNKNDLEKRSK 469
DB 448 CPAGNIIGLVGIDQPLTLTGTLTSETAHNMKVMKFSV-SPVQVAVBRKANNDLPKQVE 506
QY 470 GIGRTREDPTFKYYPTEKKEVYISGMGELHLEIYARLEREY-GCPCTIGKPKVAFRE 528
DB 507 GLKRLSKDPCVLTWM-SBSGHEIYAGGELHLEICLDLLEHDHAGVPLKISPPVAYARE 565
QY 539 TITAPVPDFTHKQSGAGQYKVIQVLEPLDPEDYTKLR----- 569
DB 566 IYES-----ESSQATALSKSPKKNRIYLAEPIDDEVSLAIENGINKPRDDFKARARIMAD 621
QY 570 -----FSDTEPGSNIKQFVPAVE-----KFLDACEKGPLSGHKL 605
DB 622 DYGMVDVTARKIWCFGPDGNGPNLVIDQTKAVQYVLAHRIKDSVVAAFQWATKGPITGEBM 681
```

```

Oy 606 SGLRFLVLOGGAHHWYDSNBSIFPLRAGEGALUKALANATC-----1LEPIMAVEYVA 657
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 RSVRVNIDVTLH-----ADAIHRGGGOTIPIPMRAITTAAGFLADPKIOEVPFLVEIOC 735
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 658 PNEFGOVIAGINRRHHGVITGOD--GVEDFYTLVADVPLNDFGXYSTELRSGTEGKGEY 714
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 726 PEQAAGGIYSVLNKKRGQGVSEQRGT-PLFTVAKYLVNBSFGFTGELRATGTGGAPE 794
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 715 TMEYSRY 721
      : : :
Db 795 QMVPDHW 801

RESULT 5
US-09-293-549-4
; Sequence 4, Application US/09293549
; Patent No. 6440409
; GENERAL INFORMATION:
; APPLICANT: G. Todd Milne
; APPLICANT: Gerald Fink
; TITLE OF INVENTION: A METHOD FOR EFFICIENT AND HIGHLY
; TITLE OF INVENTION: SELECTIVE CONTROL OF MICROORGANISMS
; FILE REFERENCE: 50078/008002
; CURRENT APPLICATION NUMBER: US/09/293,549
; CURRENT FILING DATE: 1999-04-16
; EARLIER APPLICATION NUMBER: 60/082,089
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FaSTSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-293-549-4

```





Matches 150; Conservative 78; Mismatches 181; Indels 97; Gaps 15;

```

QY 44 EKIRNIGISAHDGKTTLLTERVLYTGRAKHEVKG-KDGGAVWDSMELEKRGIT 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 ESKRTAIIISHPAGKTTTLTKLTSFGALREAGYKCKTGKTFASIDMKVQEGISV 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 QSAATFTMKDVNINIIDPGHVDFTIEVERALRVLDGAVLVCAVGVOCOTMTVNRQM 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 TSSVMQFDVDYKINILDFPGHEDFSEDYRTLVMAVDSAMVLDCAKGIETPQLKFKVC 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 KRVNPFLLPINKLDMGNSPARALQOMSKLNHTAFMOIPMGLGNFKGIYD----- 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 KMGKIFFTFINKLDMVGKPEFELDEIETLINDYPMWPMVGOMQNFPGIIDRSKTI 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 -----LIBERAIYFDGDFSOIVRGEIPAELEAATDHRQELIECVAN 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 EPPRDEBNLHNEDELKEEHALKNDSAFEQAI-----EEMLV----- 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 SDEQLGEMPLEEKIPIISDLKLAIRBATLKRSFTPVFLSGALKKGVOPLLDAVEYLPN 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 222 --DEAGAFDNE-----ALLNGELTPVFGSALANFGVQNFANVVDHAM 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 PSFVQVYALINKKDSKEKTKIIMSSRHNSHP-----FVGLAPLEV-----RFGQLTV 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 PNARQ-----TKEEVDV-----SPFDIDFGSFTFKIQANMDPKHRIRIAM 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 RSVQGLKKGDTIYNTRTRKRVRLARHADMEASTEEVYAGDICALFGI--DCASGD 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 307 RVVSGAFERMDTTLORINKKQKITSTSPADCKETVNHAV-AGDILGYDTGNQIQIG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 TPTDKANSGLSMESI-HVPDPVISIAMKPSKNLDEK--FSKGIQGFREDPTFKVYPT 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 366 TLV-----GGKQKVSFQELPQFTPEIFMKVSAKVMQKGFHKGIEQLVQEG-AIQYKTL 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 488 ENKETVYSKGELHLEIYAORLEREY 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 421 HTNQIILGAVGQLQFEVFEHRMNEY 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

# RESULT 9

```

US-08-984-618-9
; Sequence 9, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

```

INFORMATION FOR SEQ ID NO: 9:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Haemophylus influenzae

```

US-08-984-618-9

Query Match 11.7%; Score 456.5; DB 4; Length 527;

Best Local Similarity 28.7%; Pred. No. 2.6e-36;

Matches 144; Conservative 82; Mismatches 211; Indels 65; Gaps 14;

```

QY 45 KIRNIGISAHDGKTTLLTERVLYTGRAKHEVKGKDVG-AAMDSELEKRGITQ 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 10 KRTTALISHPAGKTTTLTKLTSFGALREAGYKCKTGKTFASIDMKVQEGISV 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 SAATFTMKDVNINIIDPGHVDFTIEVERALRVLDGAVLVCAVGVOCOTMTVNRQM 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 TSSVMQFDVDYKINILDFPGHEDFSEDYRTLVMAVDSAMVLDCAKGIETPQLKFKVC 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 RYNVPFLPINKLDMGNSPARALQOMSKLNHTAFMOIPMGLGNFKGIYDLI-EERA 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 LADTPIIFPMNKLDRIDRPIELDBEVNLIKRCAPITWPIGCKKLFQGVYHLAKDETY 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 IYFDGDFSOI---VRGEIPAELEAATD-----HRQELIECVANSDEQLGEMFLEE 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 190 LYQSGGSGTIOAVRVYKGNLDELVAAGDDLAAQLREBELVQASNEFEQ----- 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 KIPSIDKLAIIRBATLKRSFTPVFLSGALKKGVOPLLDAVEYLPNSEVQVYALINK 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 -----DAFIKELTPVFGTRLNFGVDHFDLGTQAPKQSQ--ADRT 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 KDSKSKTKIIMSSRHNSHPVGLAPLEV-----RFGQLTVVSVQGLKKGDTIYNT 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 287 VESAEEK-----FSGFVFKIQANMDPKHRIRVAFMVRVSGKYEKMKLKV 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 RTKRVRL-QRLARHADMEASTEEVYAGDICALFGIDCAS-GPTFDKANSGLSMESI 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 333 RIGKDVVISDALTFNAGD--RAHAEVAYAGDIIIGHNHTQIGDTFT---QGETLKFT 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 HVPDPVISIAMKPSKNLDEK--EFSKGIQGFREDPTFKVYPTDENKETVYSKGELHLE 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 387 GIPNAPBELFRIRLKDPLKQKQLKGLVQLS-EEGAVQVFRPLNNDIIVAGVGLORD 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 504 IYAORLEREYGCPCITGKPKVA 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 446 VVVSRLKTEYVNEALYENNVVA 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

# RESULT 10

```

US-08-984-618-7
; Sequence 7, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,618  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-213 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: *Staphylococcus aureus*  
US-08-984-618-7

Query Match 11.4%; Score 445.5; DB 4; Length 517;  
Best Local Similarity 28.0%; Pred. No. 3.1e-35;  
Matches 143; Conservative 78; Mismatches 179; Indels 111; Gaps 19;  
QY 44 EKIRNIGISAHIDSGKTLTERVLYYTGRIAKHVEYKGDVG-AVWDSMELRQGIT 102  
DB 8 EARKTPAIIISHPDAGKTTITLTKLVYSGAIRAGYTKGKLVNLRVTMMKVEGRGISV 67  
QY 103 QSAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLV-CAVGVQCCQMTVNRQ 161  
DB 68 TSSVQFPDVIDDLTPGHEDF--EDYRTLMAVDSAVMYIDCAKQ---VEPLTFPV 122  
QY 162 MKRVNVPFLTFINKLDRMGSNPBALQOMRSKLNHTAFNOIPMGLEGNFKGIVD----- 216  
DB 123 CMKRGITPTFINKLDRVGEKEPELIDEETLNTETTMNMPFGQSGFGLIDRKSKT 182  
QY 217 -----LIERAIFYDGFQGIYVYGBIPAEYLRMAATDRHOELIECYA 258  
DB 183 IEPFDEENILHNDPFEEDHAIITNDSDFQAI-----BELMLV-- 223  
QY 239 NSDEQGEWFLEBKIPSIDKLARATLKRSFTFVPLGSAKKNKGVOPLLDAVLEY-- 316  
DB 224 ---EEGGEAFDND-----ALLSGDLTFVFPFGSALANFGVQNFNAVVDFAF 266  
QY 317 LPNPEVQVYAILNKKDKSEKTKILMNSSRNHSHPVGLAFPLEVG---RFGQLTYVR 372  
DB 267 MKNACQTKNNVNSPFDSDS-----FSGFIFRQIANNMPKRRDRIAFKR 309  
QY 373 SYQGEIKK-GDTIYNTTRKKVRLQRLAMHA-----DWMKEASTEVEVADICALFGI-D 425  
DB 310 VVSGAFERVWMLCNVLKSK-----RSHVQRHMQITIKLVNNAVAGDIIGLYDTGN 362  
QY 426 CASGDTFTKANSGLSWESIH-VPRPVISIAMKPSKNKDLK--FSKIGIGRFTREDPTK 482  
DB 363 VOIGGTLV---GSKQTVAFODLPQTPREIFMKVSAKNVWKKQNHFKIGIEQLVQEG-AIQ 417  
QY 483 VPEFTEKNTVTSNGGELHLEIYAORLEBY 513  
DB 418 YKTLHTNQIILGAVGQLQFVFEHMKNEY 448

RESULT 11  
US-08-984-618-10  
Sequence 10, Application US/08984618  
Patent No. 6251647  
GENERAL INFORMATION:  
APPLICANT: de Lencastre, Herminia

APPLICANT: Tomasz, Alexander  
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,618  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-213 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: *E. coli*  
US-08-984-618-10

Query Match 11.3%; Score 441; DB 4; Length 529;  
Best Local Similarity 28.2%; Pred. No. 9.1e-35;  
Matches 147; Conservative 89; Mismatches 222; Indels 64; Gaps 15;  
QY 45 KIRNIGISAHIDSGKTLTERVLYYTGRIAKHVEYK--DVGAMDSMELRQGITQ 103  
DB 12 KRRTPAIIISHPDAGKTTITLTKLVYSGAIRAGYTKGKSNQHAQSDMWEKQGISIT 71  
QY 104 SAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLV-CAVGVQCCQMTVNRQWK 163  
DB 72 TSVQFPDVIDDLTPGHEDFSDYRTLTAVDCLLMYIDAKGVBDRTKLMETVR 131  
QY 164 RYNVPFLTFINKLDRMGSNPBALQOMRSKLNHTAFNOIPMGLEGNFKGIVDLIERAI 223  
DB 132 LRDPTILTFMKNLDRIDIDPMLDEVENELKIGCAPITWPIGCGKLPKGVNHLKXDety 191  
QY 224 YFD---GDFQIVR--GEIPAEYLA-----ATDRQELIECVANSDEQGEWFLEBK 273  
DB 192 LYQSGKHGTLDEVRIVKLNPNPDLDAAVGEDLAQQLRLELVKASBNPDEFLAGET 251  
QY 274 PSISDLKLARATLKRSFTFVPLGSAKKNKGVOPLLDAVLEYLPNPEVQVYAILNKD 333  
DB 252 -----TVVFGTLAGNFGVDMGLGVEMARF-----MPRQT 284  
QY 334 DSKERTKILMNSSRNHSHPVGLAFPLEVG---RFGQLTYVRSYQGEIKKGDITYNTRT 389  
DB 285 DTR-----TVEASEDKFTGFVKIQQANNDPKRRDRIAFKRVVSGYKMKMKLRQVAT 336  
QY 390 KKKVRL-QRLAMHADMMEASTEVEVADICALRIGDAS--GPFTEKANSGLSWESIHV 447  
DB 337 AKDVVISALTFPMAD--RSHVEAVYPGDIIGLHNHGTTIGIDFTT---QSEMKKFTGI 390  
QY 448 PDPVISIAMKPSKNKDL--EKFSGKIGIGRFTREDPTFKVYPTENKETVTSNGGELHLEIY 505

Db 391 PNPAPLFRRLKDPKQOLKGLVQJLS-EGAVGVFPRIISNDLIVAGVVLQDDV 449  
QY 506 AQLERREYGCPCITGKPKVA---FRETITAPVPDFPHKQOS 544  
Db 450 VARLKSEYNEAVESYNVATARWECADAKKEEFKRNES 491

RESULT 12  
US-08-984-618-8  
Sequence 8, Application US/08984618  
Patent No. 6251647  
GENERAL INFORMATION:  
APPLICANT: de Lencastre, Herminia  
APPLICANT: Tomasz, Alexander  
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
NUMBER OF SEQUENCES: 17  
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,618  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-213 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 531 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: *Dichelobacter (Bacteroides) nodosus*  
US-08-984-618-8

Query Match 11.3%; Score 440; DB 4; Length 531;  
Best Local Similarity 25.8%; Pred. No. 1,1e-34;  
Matches 143; Conservative 101; Mismatches 236; Indels 74; Gaps 18;

QY 47 RNIGISAHIDSGKTLTERVLYTGRIAKMEYKGDGV-GAVMDSMELEROGITIQ 105  
Db 13 RTEAIIISHPDAGKTTLEKLLFGAIALAGAVKGRVAHARSDDMKKEOERGISTVS 72  
QY 106 ATEPMKDVNINIIDPFGHVDFTIEVERALRVLDGAVLVLCVAVGVCOTMTVNRQMKR 165  
Db 73 VMOFPFHGKVINILDPFGHEDFSEDYRITLTAVDASLMTVDCAKGVBEERTIKLMEVCRLR 132  
QY 166 NVBELFTFINKLDRMGSPARALOQMSKLNHTAFMOIPWGLEGNFGIYDLIEBRAIYF 225  
Db 133 TTPIFFVKNLDDGDGEPWELDEIERVLIHICAPVTWPIGMGRSLKGIYHLARD-TYVF 191  
QY 226 --DGDSQIVRYE-----IPAELEAAATDHRQELIECVANSDEQLGEMFLEEKIP 274  
Db 192 YTTGKGASINHGETVVGLDNPRDLTLPDIIDDFREEI-----HFLREGVNPFDEH--- 243

QY 275 SISDLKAIIRATLKRSFPTVFGSALKKKGYOPLDAVLEYLPNPSEVONYAILINKKD 334  
Db 244 -----AYLKGELTPVYFGSALSNFVBEMLTPDAQALAPPRPRRT-----TERV 288  
QY 335 SKEKTIILNSSRHNHPVGLAFLPEVG---RFGQLTYVRSYQGLKGGDTIYNTTR 390  
Db 289 APOEKL-----TGFVFKIQANMDLKHDRIRIAPMNVNSGTFRAGKMLQVRLG 336  
QY 391 KAYRL-QRLARHADMEASTEVEVAGDICALFGIDCAS-GDPTDDKANSGLSMESINHP 448  
Db 337 REVKIPDALTPLAAREHA--QEAFAGDITGIHNGTIRIGDTPTE---GSLDPTGP 390  
QY 449 D--PVISIAMKPSNNKNDLEKFSKIGRFTREDPT--FKYFPTEKNETVISCMEILHEI 504  
Db 391 DPAPELFRVQKDPYKAKKALLGLAQCEBATOFPKRLI---GSDILIGAVLQFEV 447  
QY 505 VAQRLERREYGCPCITGKPKVAFRETTTAP--VPPDFTHKQSGAGQYGVLEPLD 561  
Db 448 VOQRLTEYVAKQFESVAVATARIIEAPNDKALQFIDKQANLHHDHYEQLYIAP-- 505  
QY 562 PEDYTKLERSDET 575  
Db 506 --SRVNLQTOERF 517

RESULT 13  
US-09-134-001C-3390  
Sequence 3390, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3390  
LENGTH: 634  
TYPE: PRP  
ORGANISM: *Staphylococcus epidermidis*  
US-09-134-001C-3390

Query Match 10.6%; Score 414; DB 4; Length 634;  
Best Local Similarity 21.8%; Pred. No. 6e-32;  
Matches 154; Conservative 91; Mismatches 202; Indels 258; Gaps 18;

QY 44 EKIRNIGISAHIDSGKTLTERVLYTGRIAKMEYKGDGVGAVMDSMELEROGITIQ 103  
Db 25 EIVRNIAIIAHVDHGKTLVLDLQSGIFRENEHVDER-----AMDSNDLEREGITTL 79  
QY 104 SAATFMMKDVNINIIDPFGHVDFTIEVERALRVLDGAVLVLCVAVGVCOTMTVNRQMK 163  
Db 80 AKCTAIDVYGTINILDPFGHADFGSEVERIKMVDGVVLVDAYEGTIPQRFVLKRL 139  
QY 164 RYVBEFLFTFINKLDRMGSPARALOQMSKLNHTAFMOIPWGLEGNFGIYDLIEBRAI 223  
Db 140 EQNLKRVVYVNNKIDRPAAP-----EGVDEVLDL----- 169  
QY 224 YRDGDSQIVRYGEIPAELEAAATDHRQELIECVANSDEQLGEMFLEEKIPISDLKAI 283  
Db 170 -----FIELBAN-DEQL----- 180  
QY 284 RRAITLKRSFTYF-----LGSALKKKGYOPLDAVLEYLPNPSEVONYAILINKKD 334  
Db 181 -----DFPVVYASAVNGTASLDBSEKQDENNQSIYETIIDVPAP-----VDN 222  
QY 335 SKEKTIILNSSRHNHPVGLAFLPEVGRFQLTYVRSYQGLKGGDTIYNTTRTKKVR 394

Db 223 SBEPLQFQALADYNY-----VGRIG--VGRVFRGKRGVDNVSLKLDGTVK 269  
Qy 395 LORLAM--HADMEASTEVEVYAGDICALFGI--DCASGDTPTDKAVSGLSMESIHVPDV 451  
Db 270 NRRVTKIFGVYGLKREIEBAQNDLIANSMDIVGETVTPHHRD--PLPLVRLIDEPT 328  
Qy 452 ISIAMPSKN-----DLEKFSKIGRFTREDPTFKVYPTENKETVLSGMEHL 502  
Db 329 LEWTFKVNNSPFAREGDYVTAQIOERLDQLETVSLKVTPTDQPSDVVAVAGREHL 388  
Qy 503 ELYAQLREYGCPCITGKPKVAFRETTAPVPFDTTHKKQSGAGQYGVIGVLEPLP 562  
Db 389 SLIEMWRE--GELOVSKRQVILRE-----IDGVLS----- 419  
Qy 563 EDYTKLEFSDFTFGSNIPKQFVPAVEKFLDACCKPGLSGHKLGLRFLVLDGAHHMVS 622  
Db 420 ----- 419  
Qy 623 NEISPIRAGEGALKALANATLCILEPIMAVEVAPNEROGVYAGINRRHG-----VITG 678  
Db 420 -----EPFERVQCEVSENAAGVIESLGARKGEMLDMMTT 454  
Qy 679 QGVEDYFTLYADVPLNDMFGSTELRSCTEGKGETTMEYSRQP 723  
Db 455 DNGLT---RLIFWVPARGMIGYTTETEMSWTRGYGINHTFEERF 496

RESULT 14  
US-09-134-001C-3574  
Sequence 3574, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3574  
LENGTH: 531  
TYPE: PRF  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3574

Query Match 9.4%; Score 367, DB 4, Length 531,  
Best Local Similarity 20.3%, Pred. No. 2e-27,  
Matches 137, Conservative 99, Mismatches 192, Indels 248, Gaps 19;

Qy 44 EKIRNIGISAHIDSGKTTLTERLVYYTGRIAKMHEVKGKGVGAVVDSMELEKRGCTTQ 103  
Db 45 ENIRNFSIIAHIDHGKSTLADRILENT-----KSVETREMQQLLDSMDLEBERGTTIK 98  
Qy 104 SAATPTMM--KD---VNIIIDTPGVADPTIEVERLRLVLDGAVLYLCAVGVCQCTMTV 158  
Db 99 LNAVRLKYAKQGETYTFHLIDPGHVDPTYESRSLACEGAILVDDAOGIEAQTLAN 158  
Qy 159 NROMKRYNVPPLTFINKLDRMGSNPARALQOMRSKLNHTAFQIQMGEGNFKGVLDI 218  
Db 159 VTLADNDIELPYNKKIDPAAEPKRVQ----- 189  
Qy 219 EERRAIVFDIDFSGIVRYGEIPALRAAATDHRQELIECVANSDEQUGEMLEKTPSISD 278  
Db 190 -----LEDVIGIDQE----- 199  
Qy 279 LKLAIRRALTKSFTFVPLGSAKNGVOPLDVAULVEYFNSEVUNYAILNKKDSKEX 338  
Db 200 -----DVLASAKSNIGIEILKIVDVVAFD-----GDPKAP 233

Qy 339 TKILNSSKHNHPFVGLAFPLEVGRFGQLYVRSYOGELKGDITVNTTRKKVYLR 398  
Db 234 LKALIFDSEYD--PYRGV-----ISSIRIIDGVVYKAGDHIKMMATGEPEVTEV 280  
Qy 399 AMHADMEASTEVEVYAGDICALFGI--DCASGDTPT-----DKANSLSMESIHV 447  
Db 281 GINTKQL--PYEEITLVGVYIILASIKVNDSDRVGDTITTLAERPADKPLQGYK 333  
Qy 448 PDPVSIAMKPSKNNDLEKFSKIGRFTREDPTFKVYPTENKETVLSG-----NGELHL 502  
Db 334 KMPWPCGIFPINDKQVNDLREALKQLNDLSLE--FEPSSQALGFRYRTGFGMLM 391  
Qy 503 ELYAQLREYGCPCITGKPKVAFRETTAPVPFDTTHKKQSGAGQYGVIGVLEPLP 562  
Db 392 EIQRIEREFIEILATAPSVIY----- 416  
Qy 563 EDYTKLEFSDFTFGSNIPKQFVPAVEKFLDACCKPGLSGHKLGLRFLVLDGAHHMVS 622  
Db 417 -----C-----ILKDG----- 422  
Qy 623 NEISPIRAGEGALKALANATLCILEPIMAVEVAPNEROGVYAGINRRHGVTGQGV 682  
Db 423 SEVSDNPAQWERKXIEH---IYEPVKATMVPNDVGAVMELCQRKQQTINMDYL 478  
Qy 683 EDY--FTLYADVPLNDM 697  
Db 479 DDIRVNIYHETPLSEV 494

RESULT 15  
US-09-095-855-207  
Sequence 207, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Rose  
TITLE OF INVENTION: Compounds and Methods for  
TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleach, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:

```

; INFORMATION FOR SEQ ID NO: 207:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-095-855-207
    
```

```

Query Match      8.7%; Score 338.5; DB 4; Length 173;
Best Local Similarity 39.7%; Pred. No. 1.9e-25;
Matches 69; Conservative 37; Mismatches 65; Indels 3; Gaps 1;
    
```

```

QY 569 EFSDETFGSNIPIKQFVPAVEKGFIDACEKGPLSGHKLGLRFVLDGAHHMVDNSNEISFI 628
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3 EFENKVTGGRIPREYIPSVNAGAQDAMQYGLAGYPLVNVKLTLLDGAHYEVDSSSEMAFK 62
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 629 RAGEGALKQALANATLCLILEPMAVEVAPNEFQGOVIAGINRRHGVITGQDGVEDYFTL 688
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 63 VAGSQVMKKAQAQAQPIILEPVMAVEVTPEQYMGEGVIGDLNRRGQIQAMEERSGARVY 122
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 689 YADVPLNDMFGSTELRSCTEGEGEYTMESRYQPCLPSTQEDVINKYLEATGQ 742
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 123 KAQVPLESMFGYVDLRSKTQGRANYSMPDSTAEVPAVNSKEIIAK--ATGQ 173
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
    
```

Search completed: July 14, 2003, 18:24:35  
 Job time : 12.4425 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 17.6298 Seconds  
(without alignments)  
4965.877 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898  
Sequence: 1 MRLGAAVVAALGRAPAS.....INKYLEATGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUECOMB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUECOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUECOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US05\_PUECOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUECOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUECOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3898	100.0	752	9	US-09-815-379-6
2	1489	38.2	692	10	US-09-815-242-11568
3	1454	37.3	700	10	US-09-815-242-11060
4	1427	36.6	704	10	US-09-815-020-248
5	1427	36.6	704	10	US-09-815-242-10363
6	1412	36.2	705	9	US-09-738-626-4055
7	1409.5	36.2	709	9	US-10-156-761-12453
8	1374	35.2	702	10	US-09-815-242-11858
9	1374	35.2	704	10	US-09-815-242-13965
10	1352.5	34.7	693	10	US-09-815-242-4877
11	1352.5	34.7	693	10	US-09-815-242-10481
12	1352.5	34.7	693	10	US-09-815-242-12333
13	1351.5	34.7	715	10	US-09-815-242-12443
14	1351.5	34.7	715	10	US-09-815-242-12736
15	1338	34.3	693	10	US-09-815-242-5238
16	1173	30.1	696	9	US-10-156-761-9338
17	667	17.1	732	9	US-10-156-761-14353
18	543.5	13.9	663	9	US-10-156-761-15035
19	516	13.2	844	9	US-10-108-605-185

20	505	13.0	845	9	US-09-991-436-110	Sequence 110, App
21	505	13.0	845	10	US-09-874-923-110	Sequence 110, App
22	478.5	12.3	884	9	US-10-128-714-4037	Sequence 4037, App
23	474.5	12.2	520	10	US-09-815-242-5735	Sequence 5735, App
24	436.5	11.2	599	10	US-09-815-242-11326	Sequence 11326, App
25	417.5	10.7	1087	9	US-10-128-714-8410	Sequence 8410, App
26	417	10.7	1013	9	US-10-128-714-3410	Sequence 3410, App
27	416	10.7	601	10	US-09-815-242-12338	Sequence 12338, App
28	416	10.7	604	10	US-09-815-242-5260	Sequence 5260, App
29	401.5	10.3	591	10	US-09-815-242-10419	Sequence 10419, App
30	396	10.2	557	10	US-09-815-242-13521	Sequence 13521, App
31	393.5	10.1	549	10	US-09-738-626-4536	Sequence 4536, App
32	393.5	10.1	607	10	US-09-815-242-13791	Sequence 13791, App
33	392.5	10.1	616	10	US-09-815-242-11136	Sequence 11136, App
34	387.5	9.9	532	10	US-09-815-242-11727	Sequence 11727, App
35	383	9.8	602	10	US-09-841-132-495	Sequence 495, App
36	381	9.8	602	10	US-09-841-132-565	Sequence 565, App
37	377	9.7	622	9	US-10-156-761-13093	Sequence 13093, App
38	368	9.4	635	9	US-10-156-761-10714	Sequence 10714, App
39	361.5	9.3	602	10	US-09-815-242-10802	Sequence 10802, App
40	353	9.1	605	10	US-09-815-242-12087	Sequence 12087, App
41	349	9.0	637	9	US-09-738-626-4721	Sequence 4721, App
42	338.5	8.7	173	9	US-10-051-643-207	Sequence 207, App
43	327	8.4	615	9	US-09-738-626-6067	Sequence 6067, App
44	325	7.1	144	9	US-09-895-913A-156	Sequence 156, App
45	217	5.6	408	10	US-09-845-335-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-815-379-6  
; Sequence 6, Application US/09815379  
; Publication No. US20030073613A1

; GENERAL INFORMATION:  
; APPLICANT: RASTELLI, LUCA  
; TITLE OF INVENTION: GERITSEN, MARY  
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: 10716/35  
; CURRENT APPLICATION NUMBER: US/09/815.379  
; CURRENT FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/191,134  
; PRIOR FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 752  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-815-379-6

Query Match 100.0%; Score 3898; DB 9; Length 752;  
Best Local Similarity 100.0%; Pred. No. 6; 7e-317;  
Matches 752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRLGAAVVAALGRAPASLGMORQVNMWACRWSGVLPEKINIGISAHIDSQRT	60
DB	1	MRLGAAVVAALGRAPASLGMORQVNMWACRWSGVLPEKINIGISAHIDSQRT	60
QY	61	TTTEVLYVTGRIAGMEHVKDGAGVAVDSMELEROGITQSATFTMMQVNIITD	120
DB	61	TTTEVLYVTGRIAGMEHVKDGAGVAVDSMELEROGITQSATFTMMQVNIITD	120
QY	121	TFGHVDFTEVERALRVLDGAVLYCAVGVQCQTMTVNRQKRYNVPFLTFINKLDRNG	180
DB	121	TFGHVDFTEVERALRVLDGAVLYCAVGVQCQTMTVNRQKRYNVPFLTFINKLDRNG	180
QY	181	SNRPALQOMSKLNHNTPAQIIPMGLEGNFRGIVDLIEBAIYVDPGSPQIVRGELPA	240
DB	181	SNRPALQOMSKLNHNTPAQIIPMGLEGNFRGIVDLIEBAIYVDPGSPQIVRGELPA	240

```
QY 241 ELRAAATDHRQELIECVANSDEOLGEMFLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEOLGEMFLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
QY 301 LKNGGVQPLDVLAVLELVPNSEVQNTAILNKQDSKERTKILMNSSRHNSHPVGLAPFL 360
DB 301 LKNGGVQPLDVLAVLELVPNSEVQNTAILNKQDSKERTKILMNSSRHNSHPVGLAPFL 360
QY 361 EVGRFQGLTVVRSYQSEELKKGDTTYNTRTRKTRLRRLARMAHMDMEASTEEVYAGDICA 420
DB 361 EVGRFQGLTVVRSYQSEELKKGDTTYNTRTRKTRLRRLARMAHMDMEASTEEVYAGDICA 420
QY 421 LFGIDCASGDTFTDKANSGLSMESIHPDPVISANKPSNKDLEKRSKIGRFTREDPT 480
DB 421 LFGIDCASGDTFTDKANSGLSMESIHPDPVISANKPSNKDLEKRSKIGRFTREDPT 480
QY 481 FKTYFTENKETIYISGGEHLHEIYQRLEREYGCCTIGKPKVAARETTTAVPDPFTN 540
DB 481 FKTYFTENKETIYISGGEHLHEIYQRLEREYGCCTIGKPKVAARETTTAVPDPFTN 540
QY 541 KKGSGAGGVGYIGVLEPLDPEDYTKLEFSDPTPSNIPKQVPAVEKGFLDACEKGPL 600
DB 541 KKGSGAGGVGYIGVLEPLDPEDYTKLEFSDPTPSNIPKQVPAVEKGFLDACEKGPL 600
QY 601 SGHKLSGLRFLVLDGAAHMDVDSNEISFIRAGEBALQALANATLCTLEPIMAVEVAPNE 660
DB 601 SGHKLSGLRFLVLDGAAHMDVDSNEISFIRAGEBALQALANATLCTLEPIMAVEVAPNE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLVADVPLNDMFGYSTLRCTBGKGEYMEYSR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLVADVPLNDMFGYSTLRCTBGKGEYMEYSR 720
QY 721 YQCLPSTQEDVINKYLEATGOLPVKKGKAKN 752
DB 721 YQCLPSTQEDVINKYLEATGOLPVKKGKAKN 752

RESULT 2
US-09-815-242-11568
/ Sequence 11568, Application US/09815242
/ Patent No. US2002061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haseelbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011a
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: PaacSeq for Windows Version 4.0
/ SEQ ID NO 11568
/ LENGTH: 692
```

```
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-815-242-11568

Query Match 38.2%; Score 1489; DB 10; Length 692;
Best Local Similarity 45.7%; Pred. No. 1.3e-115;
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;

QY 42 PNEKIRNIGISAHISGKTTLTERLVYTTGRARAKHEVYKQGVANVDSMELEROGIT 101
DB 6 PLNRIRNIGIAHIDAGKTTTSEIRLFTTGVSHKIGEV-HDG-AAATMDMEQERGIT 62
QY 102 ICSAATFTMMKDVNINIIDTPGHVDTFIEVERBALVLDGAVLVLCAGVGQCQTMTVNRQ 161
DB 63 ILSAATTCWKHQHNLIDTPGHVDTFIEVERSMVLDAVAVFCSVGQVQSEFTVRQ 122
QY 162 MKRYVPPFTFINKLDRKGSNPARLLOQMSLANTNTAPMOJPMLEGNFKIVLLIER 221
DB 123 ANKYGVPRIVFVNKDRIGANFNVENOIKORLKANPVPINIPIGAEDTFIGVIDLVQMK 182
QY 222 AIFPDGDSQIARYG-SIPAELEAATDHRQELIECVANSDEOLGEMFLEEKIPISIDL 279
DB 183 AIVANNB-TMGAKTYVEBIPSDLEKAKOYREKLEVAABQDBALMEKTLGSEEDIERI 241
QY 280 KLAIRATLKRSFTFVFLGSA LKNGGVQPLDVLAVLELVPNSEVQNTAILNKQDSKERT 339
DB 242 KGIKTCGLNMFVFLGSSSFKNKGVOQLDVAVIDYLPAPTEVVDI-KGIDPKTEE 297
QY 340 KILMNSRHNSHPVGLAFLPEVGRF-GQITVRSYQSEELKKGDTTYNTRTRKTRLRQL 398
DB 298 EVFVVS--DDEEFAGLAFKIMTDPFVQQLFVRVYRGLESGSYVNSTKDKERVGL 355
QY 399 AMHMDMEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPVISANK 457
DB 356 LKMSNKE-DIKEYIYAGEICAFVGLKDTLTDTLDCDKNA-VUERMEFPFPVHIABV 413
QY 458 PSNKDLEKFSKIGRFTREDPTFKVYFTENKETIYISGGEHLHEIYQRLEREYGCPC 517
DB 414 PKTKADQEMGVVALKLABEDPSFRVMTQEBTGOTLIGMGEHLHEIYDRLKREFKVA 473
QY 518 ITGKRVAFRETTTAVPDPFTHKQSGAGGVGYIGVLEPLDPEEDYTKLEFSDPTG 577
DB 474 EIGQPOVAFRETTIRSVSKHKYAKOSGGRGVGHVFTKLEKKEG--SGYFVNEISG 531
QY 578 NIPKQVPAVEKGFLDACEKGPLSGHKLSGRFLVLDGAAHMDVDSNEISFIRAGEBALQ 637
DB 532 VIPKEIIPAVDGIQDEAMONGVLAVGYPVDRKVTLYDSYHDVDSSENAFKIAGSMATFE 591
QY 638 ALANATLCTLEPIMAVEVAPNEFOGOVYAGINRRHGVITGQDGVEDYFTLVADVPLNDM 697
DB 592 ASRAANPVLLEPMKVEVEPPEYMGDIVGLNRRRQINSMDRLGLKIVNAPVLYEM 651
QY 698 FGSTELASCTBGKGEYMEYSRYPCLPSTQEDVINK 735
DB 652 FGSTELASATQGRGTYSMEDPHYGEVPSNAKEIYEV 689

RESULT 3
US-09-815-242-11060
/ Sequence 11060, Application US/09815242
/ Patent No. US2002061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haseelbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011a
```



```

1 CURRENT APPLICATION NUMBER: US/09/815,242
2 CURRENT FILING DATE: 2001-03-21
3 PRIOR APPLICATION NUMBER: 60/191,078
4 PRIOR FILING DATE: 2000-03-21
5 PRIOR APPLICATION NUMBER: 60/206,848
6 PRIOR FILING DATE: 2000-05-23
7 PRIOR APPLICATION NUMBER: 60/207,727
8 PRIOR FILING DATE: 2000-05-26
9 PRIOR APPLICATION NUMBER: 60/242,578
10 PRIOR FILING DATE: 2000-10-23
11 PRIOR APPLICATION NUMBER: 60/253,625
12 PRIOR FILING DATE: 2000-11-27
13 PRIOR APPLICATION NUMBER: 60/257,931
14 PRIOR FILING DATE: 2000-12-22
15 PRIOR APPLICATION NUMBER: 60/269,308
16 PRIOR FILING DATE: 2001-02-16
17 NUMBER OF SEQ ID NOS: 14110
18 SOFTWARE: FastSeq for Windows Version 4.0
19 SEQ ID NO 11060
20 LENGTH: 700
21 TYPE: prt
22 ORGANISM: Haemophilus influenzae
23 JS-09-815-242-11060

```

Query Match	37.3%	Score 1454	DB 10	Length 700
Best Local Similarity	45.1%	Pred. No. 1.1e-112		
Matches 318	Conservative 144	Mismatches 235	Indels 28	Gaps 13

```

Qy 42 ENKIRINTGISAHDSGKTTTLTERLYLTGSIARKEHKGXDGXGAVMDSHELEKRGIT 101
Db 6 PIERYRNIGISAHIDAGKTTTTERLIFYTGVSHKIGEV--HDG-AAATMDMEQBEGERGIT 62
Qy 102 IQSAATFTMMDVN-----INITDTEGHVDTFEVERLALRYLDGAVLYLCAVGVGVCQ 154
Db 63 ITSAATTAFWMSGQOPPOHRIINVIDTGHADVFEYEVRSNMRVLDDGAVMVYCAVGVGPQ 122
Qy 155 TMTVNRQMKRNVNPEPLTFEINKLDRMGSPAPALQOMRSKILNHTNFAFMQIPMGLSEGNFKGI 214
Db 123 SETVMRQANKTEVPRIAFTVNMGRGAFNLAVBELKTRLGAMALPLDLPGAENPFVGV 182
Qy 215 VDLIERAI-YFDGDFSGIVRYGEIIPAEELRAATDHRQELIECVANSDEQJGEMFLEEKI 273
Db 183 VDLIKMKAINNENADQGMTFYEEVBPANQADCESEWRONLVEAALAEASEELMEKYLGED 242
Qy 274 PSIDLKLAIRATLKRSTFTPVFGSALXKNGVQPLDLATLEYLPNBEQONVYALNKKD 333
Db 243 LTBEESISALRQRYLANEIIILVTCSAKKNKGVOMLDAVVEYIPAPDIP--AKKINP 300
Qy 334 DSKERTKIANNSSRHNS--HPFVGLAFPLEYGRF-GQLTYVRSYOGELKKGDTIYNTRTR 390
Db 301 DETE-----GERHASEDEBPSSALFIKLANDPFGNLTFFRVYSGVINSQDVIANSRQ 353
Qy 391 KKVRLQRLARNHADMEANSTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHVP 449
Db 354 KREFFRGRLVQNMNANRE-EIKEVRRAGDIAAIGLADVTCGPTLC-AIDAPILLEMERPE 411
Qy 450 PVISIAMKPSKNDLEKFSKGIGRFTRDEPFKQVFDENKRETVISGMEGLHLEIYAQRL 509
Db 412 PVISVAVPRTKADQOEKRGALGRQAQDPSFRVHTDEESGETIISGGEHLHLDIYDRM 471
Qy 510 EREYGCPCITGKPVAFRETTIAPV-PPDFTHKQSGGACQYGVKIVGLEPLDPEBYTKL 568
Db 472 KREBPKEVANIKKPQVSYRETIIRTRNVNDEGKHAQSGGRGQYGHVVIDLYPLDPEG-EGY 530
Qy 569 EFSDETBSNINPKQFVPVAVKEGFLDACRKGPLSGHKSGLAFVLDGAAHNVDSNEISFTI 628
Db 531 EFWNEIKGIVPGETIPAVNDKGIOQLKSGPLAGPVAVDLGVRLHFGSYHVDSESLAFK 590
Qy 629 RAGEGALQKQALANATLCLIEPIMAVEVAAPNEFQOQYACINRRHGVTGQDVEDYFTL 688
Db 591 LAASLAFQVARSKANPVLLEPIMKVEVEBPPEYGVSDIGULSRRAAMNGGANEFPVKI 650
Qy 689 YADVPLANDMGYSTELRACSTBEKGSEITMEYSRIOYDCLPSTJQEDVAT 733

```

Db

651 YAEVPLSEMFYATDLRSQTQGRASYSMEPLKYAEAPTSVAAAVI 695

RESULT 4  
 US-09-912-020-248  
 Sequence 248, Application US/09912020  
 Patent No. US20020045592A1  
 GENERAL INFORMATION:  
 APPLICANT: Zykkind, Judith  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Trawick, John  
 APPLICANT: Forsyth, R. Allyn  
 APPLICANT: Froelich, Jamie M.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
 TITLE OF INVENTION: ESCHERICHIA COLI  
 FILE REFERENCE: EITRA.001DV1  
 CURRENT APPLICATION NUMBER: US/09/912,020  
 CURRENT FILING DATE: 2001-07-23  
 PRIOR APPLICATION NUMBER: 09/492,709  
 PRIOR FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: 60/117,405  
 PRIOR FILING DATE: 1999-01-27  
 NUMBER OF SEQ ID NOS: 485  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO 248  
 LENGTH: 704  
 TYPE: PRF  
 ORGANISM: E. COLI  
 US-09-912-020-248

Query Match	36.6%	Score 1427;	DB 10;	Length 704;
Best Local Similarity	44.2%	Pred. No. 2e-110;		

Qy	42	ENEKIRNIGISAHIDSKETTLTEBVLYYTGRIAMHEVYKQGVAMDSHELRQKIT	101
Db	6	PIAYRNIGISAHIDACKTYYTERLIFYTYGNHKGIEV--HDG-AAATMDWBOEOERGIT	62
Qy	102	IOSAATPMWKDV-----NINIDPRGVDFIETBRLRLVYDGVVLVLCANGVQOC	154
Db	63	ITSAAITAFSGMAKOYEPRRINIIDPRGVDFIETBRSNRVLVDGAMVYCAAGVQPO	122
Qy	155	TMTYNROMKRYNVPFLTFINKLDRMGSPAPALQOMSKLNHTAFQOIPMGLEGNFKGI	214
Db	123	SETYWRQANKYKUPRIAFVKNMDRMGNFLKVNQITRRLGANVPQLAGAEHFTGV	182
Qy	215	VDLIEERAI-YFDGDFSQIVRGEIPALBRLAATDHRQELIECVANSDEOIGENFLEKI	273
Db	183	VDLIVKMAKIMNWDADQVTFREYEDIPADWELANEMHQNLIESAABASBELMEKYLGBE	242
Qy	274	PSISDLKALRRATLKRSFTPVPLFGSALKKKGVQPLLDVLEBYPNSEVQONTAIIANKKD	333
Db	243	LTEAEIIGALRQRLVNEIILVYTCGSFKNKGQVAMDAVIDYPSVLVDP--AINGILD	300
Qy	334	DSKEKTIILNSSRHNS--HPVGLAPLLEVGPR-GQITVRSVQEGELKKGDDTYNTRR	390
Db	301	DGKDTP-----AERHNSDDEBPFSALAKKTIITDPFVGNLTFPRYSGVNVSGDITLANSVKA	355
Qy	391	KKVALQRLAHMADDMWEASTEBEVYACDICALFGI-DCASGDTPTDKNSGLSMESIHPD	449
Db	356	AREPFGIIVGMHANKBE-EIKETRAGDIIAAGIKDVTOTDCLD-FDAPILIERMEPE	413
Qy	450	PVSIANKPSNKUDLEKFSKGIGRFTREDPTFKYVPDTEKKEIYVSGEILHLEIYAQRL	509
Db	414	PVSIABEPKTKADQECMGALTRLAKEDBESFRWTEBESNQITIIACMGELHLDIIVDRM	473
Qy	510	EBEYGCCTIGKRVARRETTIAPV--PFDTHKKQSGAAGYGVIGVLEPLDPEBDYTK-	567
Db	474	KREBNVEANVGKQVARETIRQKYDVIDEESKAAQSGRQOYGVAVIDMPLLEGSPK	533





```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11858
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11858

```

Query Match 36.1%; Score 1406; DB 10; Length 702;

Best Local Similarity 43.6%; Pred. No. 1.1e-108; Indels 30; Gaps 13;

Matches 308; Conservative 121; Mismatches 248;

```

QY 42 PNEKINIGISAHIDSGKTTLTERVLYYTGRIAKMEHVKGKDGAVGAVMDSMELERQGIT 101
DB 6 PIELYINIGIVAVDAGKTTTERILFYGVNHKMEV--HDC-AATMDMVEQGERGIT 62
QY 102 IGSAAFTFMWKD-----VNIIITDPGHVDTTIEVERALVLDGAVLVLCVAGVQC 153
DB 63 ITSAAFTAFWGSGTKQPPHRYRNIIIDTDPGHVDTTIEVERSLVLDGAVVVFSGAGGVGP 122
QY 154 QMTVTRAKMRVAVPFLTFINKLDRMGSNPBALQOMRSKLNNTAFMOIPMGLGNFPG 213
DB 123 QSETVTRQANKIVPRILAVYKMKQADFLVVAQIKQRLGHVPPVPIQALISSENFSG 182
QY 214 IVDLIERAIYF-DGDFSQIVRYGEIPALBLAAATDHRQELIECVANSDEQLGEMPLEEK 272
DB 183 QIDLVKMAIYVMDADQGSTSYREBEIPALBLAAEMBRAMVAAAEADDELANKYLEB 242
QY 273 IPIISDLKAIIRATLKRSFTPVFLSALKKNGVQPLDAVLEYLNPSEVQNYAIALNKK 332
DB 243 ELISIEIKKGLRQRLANQIVPAVLGSSFKCKKGVPLVDAVIDYLPAPSEIP--AIRGTD 300
QY 333 DSKERTKILMNSRH--NSHPFVGLAFPLEVGRF--GOLTYVRSYOGELKKGPTIYNTRT 389
DB 301 PDDEEK-----HDERHADDEPFSALAFKIATDPFVGILTPARVYSGVLTSGDAVLSVK 355
QY 390 RKKVRLQRLARMAHDMMEASTEVEVYAGDICALFGI--DCASGDTF--TDKANSGLSMESI 446
DB 356 GKKEKRGVWQWAMNQRD--EIKEVRAGDIALALGMQDVTTGDTLCAIDKP---IILERND 411
QY 447 VDPPIVSIAMKPSNKQDLEKFSKIGQRFREDPTFRVYPTDENKETVIGSGMLHEIYA 506
DB 412 FPDPIVSIAMKPSNKQDLEKFSKIGQRFREDPTFRVYPTDENKETVIGSGMLHEIYA 471
QY 507 ORLERVYGCPCICGKRVAFRETIPTAPVPPDFTHKKOSGAGQGVAVIGVLEPLDEEDT 566
DB 472 DRRRRRFGVANIIGKQVAVRETIPTCEIEGKQVOSGGRGQFGHCWIRFPAAD--EGGE 530
QY 567 KLEFSDTEFSNIPKQFVPAVEKGLDACEKGPLSGHKLSGLFVLDGAMHNVADNEIS 626
DB 531 GLEFHNIEVGVIPREFIPIQIGIEDQNGVLAQYPLIGLKATYDSDSHYDVSEWA 590
QY 627 FTRAGGALKQALANMTCLIEPIVAVEVVAENEFQGVYAGINRHHGVITGQGVEDYF 686
DB 591 FKLAAMATKQLSQKGAVALLEPVMTVEVYTPEDYGDVMDGNRRRRLQGMEDTPAGK 650
QY 687 TLYADVPLNDMPGYSSTELRSCTEGKGEYTMESRYQPCLPSTQEDVI 733
DB 651 VTRAEVPLGEMFGYATDVRSMSQGRASYMEFVRYAEVSPASVAGIV 697

```

# RESULT 9

US-09-815-242-13965

Sequence 13965, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl U.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13965

LENGTH: 704

TYPE: PRT

ORGANISM: Salmonella typhi

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(704)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-13965

Query Match 35.2%; Score 1374; DB 10; Length 704;

Best Local Similarity 42.6%; Pred. No. 5.3e-106;

Matches 306; Conservative 126; Mismatches 250; Indels 36; Gaps 15;

```

QY 42 PNEKINIGISAHIDSGKTTLTERVLYYTG--RIAKMEHVKGKDGAVGAVMDSMELERQ 98
DB 6 PIARRNIGISAHIDAGKTTTERILFYGVNHKIGEXH----DG-AATMDMVEQGER 59
QY 99 GITTSAAFTFMWKV-----NIIITDPGHVDTTIEVERALVLDGAVLVLCVAGV 151
DB 60 GITTSAAFTAFWGSGAKQYEBRNIIIDTDPGHVDTTIEVERSMKVLGAVVYVCAVGV 119
QY 152 QOQMTVTRAKMRVAVPFLTFINKLDRMGSNPBALQOMRSKLNNTAFMOIPMGLGNF 211
DB 120 QOQSETVTRQANKIVPRILAVYKMKQADFLVVAQIKQRLGHVPPVPIQALISSENF 179
QY 212 KGIIVDLIERAI-YDGFPSQIVRYGEIPALBLAAATDHRQELIECVANSDEQLGEMPLE 270
DB 180 TGVVOLLVKKAIINMMDADQGVTFEYEDIPADWQDLANEHQLISAAASELMKEYLG 239
QY 271 EKIPISDLKAIIRATLKRSFTPVFLSALKKNGVQPLDAVLEYLNPSEVQNYAIALN 330
DB 240 GHELTEEBIKQALRQRLANNEIILVTCGSAFQKNGVQAMLDVAVIDYLPSPVDP--AING 297
QY 331 KDDDEKERTKILMNSRHNS--HPFVGLAFPLEVGRF--GOLTYVRSYOGELKKGPTIYNT 387
DB 298 IIDDGKDP-----AERHASDDEPFSALAFKIATDPFVGNTLTFRYYSIGVANSQDITVLS 352

```

Qy 388 RTTKVRLQRLARMLADMMEAESTEYVYAGDICALFGI-DCASGDTFTDKANSGLSMESI 446  
Db 353 VTAARERFERIYQMHANKEE-EIKEVAGDIAAAGLKVDITGDTLCD-PEXPILLERME 410  
Qy 447 VDDPIIISAMKSNKNDLEKPSKICRFRREDPTFRVYDTENKEVVISGMELHLEIYA 506  
Db 411 LPEPIISAXAVEBKTADXXKMGALGRKLAKEDPSFPVWDEESNQTIIAGMELHLDIIV 470  
Qy 507 ORLEREYGCPCITGKPKVAFRETITAPV-PFDTHKKOGSAGAGYGVIGVLEPLDPEDY 565  
Db 471 DMKKEFNVANVGKQVAYREAIRAKYTDIEGKAHKGSGRGQYCHVIDYIMPLEPESN 530  
Qy 566 TK-LRESDTEFGSNIPKQFVPAVEKGFUDACEKPLSGHLSGLRFLVLOGAHHWVDSN 624  
Db 531 PKGYEPIINDIKGVIPGEYIIPANVDKGIQGLKSGPLAGYPVVDLGVRLHFGSYHDVDSSE 590  
Qy 625 IGFIRAGEBALKQALANATLCLIEPIMAVEVVAPEFQGVAGINRRRGVITGDDGVSD 684  
Db 591 IAFKLAASIAFKEGFPKAKPVLLEPIMKVEETPEENTDVGIDLSRRXMKKGQSESVT 650  
Qy 685 YFTLYADVPLNDMFGYSTELASCTEGKEGYTMEYSRYQCLPSTOEDVINKYLEATGQ 742  
Db 651 GKXIHAEVYALXEMFGYAKQXRSTXGRASVTMKFLKD----XAPXVAGAVIEARGK 704

RESULT 10  
US-09-815-242-4977  
Sequence 4977, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4977  
LENGTH: 691  
TYPE: PRF  
ORGANISM: Enterococcus faecalis  
US-09-815-242-4977

Query Match 34.7%, Score 1352.5, DB 10, Length 691;  
Best Local Similarity 42.1%, Pred. No. 3, 2e-104;  
Matches 292, Conservative 111, Mismatches 256, Indels 15, Gaps 9;  
Qy 44 EKIRNIGIAHIDSGTTLTERLYTGRAKNHEVKGSDVAVNDSMELEQSGITITQ 103  
Db 8 EKIRNIGIAHIDSGTTLTERLYTGRAKNHEVKGSDVAVNDSMELEQSGITIT 64

Qy 104 SAATFTMKNDVNNIITDPGHVDTIEVERALRVLDGAVLVLCAGVYQOCOTMYNRQMK 163  
Db 65 SAATTAQKRGYRVNIITDPGHVDTIEVERALRVLDGAVLVLCAGVYQOCOTMYNRQAT 124  
Qy 164 RYNNVEPLFIKLDMSGNSPARALOQMSKLNHNTPAFNOIMGLEGNGKGVLDLEBAI 223  
Db 125 EKVYRIYFCNMKMDIGADFPYVSVELHRLQANNAHPQIPIGAEDEDTGLIDLKMAE 184  
Qy 224 YFDGDSQIVRYGEIIPALRAAATDHRQELIECVANSDEQLEGMEPLFEKIPISIDLKAI 283  
Db 165 IYTNLDGTDIGETDIPEDYLEKAGQWRKLYEVAVAETDEDLMMKYLEEBEITRELVAGI 244  
Qy 284 BRATIKRSFEPVFLGSLAKNGVQPLDVALEBYLPNPSVQNYALINKKDSKEKTKILM 343  
Db 245 RQATINVEFPVLASAPKNGVQLMLDAVLDPSPIDAIKIDITKTD-EETTR--- 300  
Qy 344 NSSRNSHPFVGLAPPLEVGRF-GQLTYTVRSYQGLKKGDITTYNTRTKVRLQRLAMH 402  
Db 301 --PADDEAPFASLAKVMTDPPVGRITFFRYSGVLESQSYVLAASKKKEKIGHILQW 358  
Qy 403 ADMMEASTEYVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPVISIAMKPSNK 461  
Db 359 ANTRQ-EIDKYSSGDIILAAVGLKQDTTDTLC-ALDAPVILLESIEFPDPVIOVAVEPSK 416  
Qy 462 NDLEKFSKIGRFTEDPTFKYFDTENKEVVISGMELHLEIYAORLEREYGCPCITGK 521  
Db 417 ADQDKGVALQGLAEEDSPFRVETNVEETGVISGMELHLDVLDRMKREKFEVANYGA 476  
Qy 522 PKVAFRETITAPVPDPFTHKKQSGAGQYGVIGVLEPLDPEDYTKLEFSDDTSPSNIPK 581  
Db 477 PQVSARETFRATKKEGKFRVQSGKQYGHWVEFTP--NEEVKGFEEVNAVIGGVPR 534  
Qy 582 QFPAVEKGFUDACEKPLSGHLSGLRFLVLOGAHHWVDSNISIFIRAGEBALKQALAN 641  
Db 535 EYIPAVEKGLDSMNGVLAGYPLVDITAKLYDSYHYDVDSHETKFRVAASALAKAAN 594  
Qy 642 ATLCLIEPIMAVEVVAPEFQGVAGINRRRGVITGDDVEDYFTLYADVPLNDMFGYS 701  
Db 595 ANPVILEPMMKVTIIVPBDYLDGIDNGHTSRRGRVEGMEAHGNSQIVNANVPLAEMFGYA 654  
Qy 702 TELASCTEGKEGYTMEYSRYQCLPSTOEDVINK 735  
Db 655 TLRASATQGRGTFMMVFPHYEDVPKSVQEBIHK 688

RESULT 11  
US-09-815-242-10481  
Sequence 10481, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

```

? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 10481
? LENGTH: 693
? TYPE: PRY
? ORGANISM: Enterococcus faecalis
?
US-09-815-242-10481
.
```

Query Match	34.7%	Score	1352.5	DB	10	Length	693
Best Local Similarity	42.1%	Pred	No. 312	DB	10		
Matches	252	Conservative	131	Mismatches	256	Indels	15
						Gaps	9

Qy	44	KKIRIRIG SAHDSCKTTL TEREVL YGTOR AKMHEV KGDG GAW DSME LEROG ITLO	103
Db	8	EKTRNIG GMAHVDACK TTTER IL YTGK HKIGST--HEG-ASQMDM EBOEQER G ITT	64
Qy	104	SAATF TMAKD VNI DTP GHVDT E E ER A L RV LDG AV L CAV G V OCQ OTM VNRQMK	163
Db	65	SAATF AQK GYR VNI DTP GHVDT E E VOR S L RV LDG AV L D S G S VEP OTET VWRQAT	124
Qy	164	RYNP E L FI NK L DM G NS P AR ALQ OMS K LN H N A P O I P MG E GN K G I VD L E E A I	223
Db	125	EYK P R V I F C NM D K I G AD F E Y S V E S L D R O A N A H P I Q I P G A E D T G I I D L K MA E	184
Qy	224	YFDG F SO I V A Y G E I P A E L R A A T H Q E L E C V N S D E O L E G E ME L E E K I P S I D L K A I	283
Db	185	IY TND L G D I O E T D I P E D Y L E K A Q W R E V A V A E T D E D L MM K Y L E G E B I T E E L V A G I	244
Qy	284	RR A L K E F T P V L G S A L K N G V O L D A V E L P N B E S V O N Y A I L N K D S K E K I L M	343
Db	245	RO A T I N V E F P L A S A F K N G V O L D A V E L P S P D I A I K I D T K T O--E E T R--	300
Qy	344	NSS R N S H P F L A P L E V E R F--G O L T V R S V O G E L K K G D R Y N T R T R K V L O L T A H H	402
Db	301	--P A D E A P F A S I A F K N W T D P V G L T F P R Y S V L S G S Y L N S K G K E R I G I L O H H	358
Qy	403	AD ME A S T E E Y A G D I C A L F G I--D C A S D T F D K A N S G L S E S I H V D P V I S I A N K P S N K	461
Db	359	AN T R O--E I D K Y S G I A A V L K D T T D T D I C--A L D A V I L E S I E P P V O I V A V E P R K S	416
Qy	462	N D L E K S G I G F T E D P F K Y P P T E K E T Y I S G E L H E I Y A Q R E E Y G C R I G	521
Db	417	A D O D K G A L O K L A E D E D S F R E T N V E Y G E I Y S G E L H D V L D R N K R E K E Y A N G A	476
Qy	522	P K V A F E T I T A V P P D F T H K Q S G A G O Y G K I G V L E P D E D Y T K L E F S D E T F G S N I P K	581
Db	477	P O V S R E F R A T K E G F R V Q S G Q G H W V E F T P--N E V G F E F E N A I V G V P R	534
Qy	582	Q P V A V E G F D A C K G P L S G H K S G L F V L O D G A H N V D N S E I F I A G E A L Q A L A N	641
Db	535	E Y I P A V E G L E B S K N G V L A G P L V D I A K I Y D S Y H O V D N S E I T F R V A S A L Q A A K N	594
Qy	642	A T L C I E P I N A V E V A P N E F O G O Y V A G I N R R G V T T G O D E V E Y T L A D V P L N D F G S	701
Db	595	A N P I L E B M K T I V P E D Y L G I N G H T S R R G R V E G N E A H N G O I V A N A P L A M F G A	654
Qy	702	T E L S C T E G K E Y T H E S R Y O P C L P S T O D V N K	735
Db	655	T T R E A T Q R G T F N V F D H Y E D V P R S V O B E I I K	688

RESULT 12  
US-09-815-242-13233

- ; APPLICANT: Haselbeck, Robert
- ; APPLICANT: Ohlsen, Karl L.
- ; APPLICANT: Zyskind, Judith W.

```

APPLICANT: Wall, Daniel
APPLICANT: Traawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1323
LENGTH: 693
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-815-242-1323

```

Query Match	34.7%;	Score 1352.5;	DB 10;	Length 693;
Best Local Similarity	41.9%;	Pred. No. 3.2e-104;		
Matches 291;	Conservative 140;	Mismatches 248;	Indels 15;	Gaps 9

```

Oy 44 EKIRNIGISAHIDSKDTTLTEBVLUYGRIANMEHVKGQVAGVYMDMSLEBORGITTO 103
Db 8 EKIRNIGIAHAVDAGKTTTERTTILUYGKIHKIEET--HSG-ASQMDMMQEOBERGTTT 64
Oy 104 SAATFTMMKDVINIITDPGHVDFTEVERALRYLDAVAVLCAVGGVOCQCTMTVNRQMK 163
Db 65 SAATTQOMNNHVNIIITDPGHVDFTEBQSLRYLDAVAVLDSQSVBEPQTETVNRQMT 124
Oy 164 RYNVPFLTEINKLDRMGSNPAPALQOMRSKLNHTAFMOJPMGLEGNFKGIVDLEBRAT 223
Db 125 EYGVPRIVFANMKDKIGADFLYVSTLHRLQANAHPIQPIGSEDDPFRGIIDLIKMAE 184
Oy 224 YFPGDSSQIVRGEI PAELBPAADHRELLECJANDEOLGEHFELEKPISTJDLKAI 283
Db 185 IYNDJGTDLBEDIPAEYLDQAOEYREBKLEAVALTEDEBEMKYLEGEBITHEELKAGI 244
Oy 284 RRATLRSEFTPELISALKNGVOPLDAYLEVLPNPSEVONVALINKKODSEKTKILM 343
Db 245 RKATINVEFPPLCSAFKNGQVQMLDAYVIDYLPSPDLIPALGINPDTDAE----IR 300
Oy 344 NSSRHNHPVGLAEPLLEVGRF-GOLTYVRSYOGELKKGDTIYNTTRKKVRLQRLARNH 402
Db 301 PAS--DEEPFALAKRIMTDPVGRITFRFYSGVLDSGSVLTNSKXKERIGRILOMH 358
Oy 359 ADMMEASTEYVAGDICALFGI-DCASGDPTTDKANGSLMESIHVDPYIATIAMKPSNK 461
Db 359 ANSRQ-EIDTVYSGIIAAVGLKDTTGDLSLTDE-KAKIILESTINVEPIQIOLMVEBPSK 416
Oy 462 NDLERSKSGIKGFTEDEPFGKYPTDETEKENVISGMBELHLEIYAQRLBEEVOCPICTG 521
Db 417 ADDDKKGIALLQCLABEDPFRVETNVEGEFVVISGMBELHLDVLDVDMRSEFVEANVGA 476
Oy 522 PKAFARETITTAVPDPDPFHKKOSGGAGYGVIVLSEPLDPEDYTKLEFDEFETGSIIPK 581
Db 477 PQVSYHETRAQTQARGFPFKQSGGKQGFV--WIEFTPNBEKGEFENALVGVVPR 554
Oy 582 QPVAPEKGFPLDACEKGPLSGHKLGLFVLQDGAHHVDSNLSIFIRAGEGALKQALAN 611

```

```

Db      535  BFLPAEKLIVSMANGVLAGYPMVAVKAKLYDGSYHDVTSSETAKIASLSLKEAAS 594
Qy      642  ATLCILEPIMAEVAVPAENFEOGOVAGINRRHGVITGODGVEDYFTLYVADVPLNDMFGS 701
Db      595  APPALIEPMMVLTITTYPERLSDVMGHVTRKGRVGMPEHAGNSQIVRAVPLAEMFGYA 654
Qy      702  TELRSGTEKGEYMEYSRYOPLPSTOEDVINK 735
Db      655  TVLRASAGRGTFMVFHDHYEDVPKSVGEIILK 688

RESULT 13
US-09-815-242-12443
; Sequence 12443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12443
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12443

Query Match      34.7%; Score 1351.5; DB 10; Length 715;
Best Local Similarity 40.3%; Pred. No. 4.1e-104;
Matches 295; Conservative 133; Mismatches 269; Indels 35; Gaps 10;

Qy      16  RAPASIGQKQKQVNMWACWSSSGVLPNKIRNIGSAHIDSKTTLTERVLYYGRIRAK 75
Db      3  RASPIVSIYFSNINWKEKTYMA-EPSEKTRNIGMAHIDGKTTTRILYYGRIRAK 61
Qy      76  MHEVKGKLVGAVMSMELEROGITIGSAATFTMKDVININIDTPGHVDFTEVERAL 135
Db      62  IGET--HEG-ASQMDMBEODRGITITSAATTAAMEGHRVINIDTPGHVDFTEVEREL 118
Qy      136  RVLDAVAVLCAVGVCQCTMTYNNRQKYNVPLFTFYKLDPMGSPAPALQOMSKUN 195
Db      119  RVLDAVAVTLDAOSGVEPQETETWRQATTYGVPRIVFVNMKDLGANSFVSSTLDRLO 178
Qy      196  HNTAFWQIQMGLGKNGKGIIVDLIEBAIFDGFSGQIVAYGEIPARLRAATDHRQEL 255
Db      179  ANAAAPQIPIGADEDEBAIIDLVEMKCFKTYNDLGEIEIEIEIFPEHLDRAEBAASLE 238
Qy      256  CVANSDEQJGEMFLEBKIPISIDLKLAIRRAATLKRSFTVFILGSALKNGVOPLLDAVLE 315
```

```

Db      239  AVALTSDEIMKCYLDEEBISSELKEAIRQATTNVEFYVPLCGTAFKXNGVQJMLDAVID 298
Qy      316  YLPNSEVO-----NVALINKDDSEKKTILNNSRHNSHFFGLAPLEVGSR 365
Db      299  YLPSPDVAKPIIIGHRASPEEERINKADDSAE-----PALAFKWTDPY 343
Qy      366  -GOLTYVNSYOGELKKGDTIYNTTRKRVRLQRLARMHADMEASTEERYAGDICALPGI 424
Db      344  VGLTFEPVYSGTWSGIVNKSSTGKRERGRLLQWANSRQ-BIDVYSGDIAAAGL 402
Qy      425  -DCASGDTFTDKANGSLNESIHVPDIYISLAKMSNODLEKSKGIGRFPREDPTFKV 483
Db      403  KDTGTDLTGCEKND-IILSEWPEPEPIHLVSEBKSADODKMTQALVKQLEBPTFPA 461
Qy      484  YFDTENKETVLSGMEHLLEIYAQLEBEYGCPTGPKVAFREITAPVPFDPTHKKQ 543
Db      462  HTBESTGVITIGWELHLIDLVDPMKKEFNVECVAGAPMVSRYRTFKSSAOGKPSRQ 521
Qy      544  SGAQOYQKVIQVLEPLDPEDYTKLFSDDETFGSI PKQFVPAVEKGFPLDACCKPLSGH 603
Db      522  SSGRGQYGDV--HIFFTNETGAGFEFENALVGVVPREYIPSVAGLKDAMENGVLAGY 579
Qy      604  KLSGLRFLVQDGAHHMVDNSNEISPIRAGEGALKOLANATLCILEPIMAEVAVPAENFQ 663
Db      580  PLIDVAKLYDGSYHDVDSSEMAFKIAASLAKKACDPIILPMMKVTIEMPEBYWG 639
Qy      664  QVIAGINRRHGVITGODGVEDYFTLYADVPLNDMGYTELRSCTEGGEYMEYSRYOP 723
Db      640  DIMGVTGRGRGVDDMEPRGNAQVNAVYPLSEMGVATSLKSNTOHGTYMYFDHYAB 699
Qy      724  CLPSTOEDVINK 735
Db      700  VPKSIADIEDILK 711

RESULT 14
US-09-815-242-12736
; Sequence 12736, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Elittra, 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12736
; LENGTH: 715
; TYPE: PRT
```

ORGANISM: Staphylococcus aureus  
US-09-815-242-12736

Query Match 34.7%; Score 1351.5; DB 10; Length 715;

Best Local Similarity 40.3%; Pred. No. 4.1e-104; Matches 295; Conservative 133; Mismatches 269; Indels 35; Gaps 10;

```
QY 16 RAPASIGMQRKQVWAKACRWSSSGVLPNEKIRNIGISAHIDSQKTLTERVLYYGRIRAK 75
DB 3 RASPIVSYLFSNNWMEKKYMAR-EFSLEKTRNIGIMAHIDAKTTERILVYGRIRIK 61
QY 76 MHEVKGKGVGAVMSNELEORQITITQSAATFMKDVNINIITDPGHVFTIEVERAL 135
DB 62 IGET--HEG-ASQMDMMEQODRIGITITSATTAEGHGVNIITDPGHVFTIEVERSL 118
QY 136 RVLDSGAVLVCAVGOCCMTNRMKRYNVEPLFINKLDMGNSPAPALQOMSKUN 195
DB 119 RVLDSGAVLVDAOSGVPEOTETWRAQATTVGVRIYVNMKDLGANFEYSVSTLDRQ 178
QY 196 HNTAFMOIPMGLEGNFGVLDLIEERAIYFDGFSQIVRGEIPAEIRAAATDROBLE 255
DB 179 ANAAPIQPIGADEFEPAITDVEKMKCFKTNLGTIEIEIEIPEDHLDRAEASLSIE 238
QY 256 CVANSDEQLEMPLEEKIPISDLKLAIKRRATLKRSFTVPFLSGALKKNGVQPLDAVLE 315
DB 239 AVAETDELMEEKYLGDEISVSELKEAIRQATTNVEFFVPLCGTAFKNGVQPLDAVID 298
QY 316 YLPNPEVQ-----NYALINKKDSKEKTKILMNSRRNSHPVGLAFLPVGRR 365
DB 299 YLPSPIDVVKPIIGHRASNPREEVIAKADDSAE-----FAALAKVMTDP 343
QY 366 -GOLTVYRSYQELKKGDTIYNTRTKVYLQRLAMHADMEASTEVEYAGDICALFGI 424
DB 344 VGLTJFPRVYSGMTSGSYKNSGTGKREKRVGLLOMHANSRQ-EIDTVYSGDIAAVGL 402
QY 425 -DCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGRFTREDPTFKY 483
DB 403 KDTGTDLTLCGEKND-IILSEMEFPPEVILHSVEPKSKADQDQMTQALVLOEEDPTFHA 461
QY 484 YPTEKKEIVYISGMGELHLEIYVQRLEREYGCRCITGKPKVAFREITTAVPDPFTHKQ 543
DB 462 HDEEYGOVITIGMGEHLHDLIVDRMKKEPNVBCNAPVSYRETFKSAQVOGKPSRQ 521
QY 544 SGAGQYGVKIVGLEPLDPEDYTKLEFSDETFCGSIPIKQFVPAVEKGFLDACEKGPLSGH 603
DB 522 SGRGQYGV--HIEFPNETGAGFEFENAIIGGVPREXIPEVEAGLDMAMENGLAGY 579
QY 604 KLSGLRFLVLDGAHMYDSNEISFIRAGBGLKQALANATLCLLEPIMAVEVVAENPEFG 663
DB 580 PLIDVAKLYDGSYHVDSEMAFKTAAALAEAAKCDPVLLEPMKVTIEMPEEYNG 639
QY 664 QVLAGINRRHGVITGDGVEDFTLYADVPLNDMPGSTELBSCTGKGEVMEYSRQ 723
DB 640 DINGDVTSRRGRVDGMEPRGNAOVNAYVPLSEMFGVATSLRSNTGRTYTWYFPHYAE 699
QY 724 CLPSTQEDVINK 735
DB 700 VPKSIAMEDIKK 711
```

RESULT 15  
US-09-815-242-5238

Sequence 5238, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haebelbeck, Robert  
APPLICANT: Onlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA 011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5238

US-09-815-242-5238

Query Match 34.3%; Score 1338; DB 10; Length 693;

Best Local Similarity 41.1%; Pred. No. 5.2e-103; Matches 289; Conservative 128; Mismatches 253; Indels 34; Gaps 9;

```
QY 44 EKIRNIGISAHIDSQKTLTERVLYYGRIRAKMHEVKGKGVGAVMSNELEORQITITQ 103
DB 8 EKTRNIGIMAHIDAKTTERILVYGRIRIKIGET--HEG-ASQMDMMEQODRIGITIT 64
QY 104 SAATFMKDVNINIITDPGHVFTIEVERALRVLDAVLVCAVGOCCMTNRMK 163
DB 65 SAATTAAMEGHVNIITDPGHVFTIEVERSLRVLDAVTVDAOSGVPEOTETWRAQT 124
QY 164 RNVNPEPLTINKLDMGNSPAPALQOMSKUNHNTAFMOIPMGLEGNFGVLDLIEERAI 223
DB 125 TVGVRIYVNMKDLGANFEYSVSTLDRLOANAAPIQPIGADEFEPAITDVEKCF 184
QY 224 YFDGFSQIVRGEIPAEIRAAATDROBLECVANSDEQLEMPLEEKIPISDLKLAI 283
DB 185 KTNLGTIEIEIEIPEDHLDRAEASLSIEAVATSELMEEKYLGDEISVSELKEAI 244
QY 284 RPAATLKRSFTVPFLSGALKKNGVQPLDAVLEYPNPEVQ-----NYALINKD 333
DB 245 RQATTNVEFFVPLCGTAFKNGVQPLDAVIDLPLSPIDVVKPIIGHRASNPREEVIAKAD 304
QY 334 DSEKTKILMNSRRNSHPVGLAFLPVGRR-GOLTVYRSYQELKKGDTIYNTRTKK 392
DB 305 DSAB-----FAALAKVMTDPYVGLTJFPRVYSGMTSGSYKNSGTGK 349
QY 393 VRLQRLAMHADMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHVPDPV 451
DB 350 ERVGRLLQMHANSRQ-EIDTVYSGDIAAVGLKDTGTLGCKEND-IILSEMEPEV 407
QY 452 ISANKPSNKDLEKFSKIGRFTREDPTFKYIPTEKKEIVYISGMGELHLEIYVQRLER 511
DB 408 IHSVEPKSKADQDQMTQALVLOEEDPTFHAHTDEETGOVITIGMGEHLHDLIVDRMK 467
QY 512 EYGCPCITGKPKVAFREITTAVPDPFTHKQSGAGQYGVKIVGLEPLDPEDYTKLES 571
DB 468 EFNVECNVGAPEVSYRETFKSAQVOGKPSRSGRQGV--HIEFPNETGAGFE 525
QY 572 DETFGSNIPIKQFVPAVEKGFLDACEKGPLSGHLSGLRFLVLDGAHMYDSNEISFIRAG 631
DB 526 NAIYGVVPREXIPEVEAGLDMAMENGLAGYPLIDVAKLYDGSYHVDSEMAFKTAA 585
QY 632 EGALKQALANATLCLLEPIMAVEVVAENPEFGQVLAGINRRHGVITGDGVEDFTLYAD 691
```



Db	586	SLALKEAAKCDPVILIEPMKVTIEMPEEYMGDIMGDVTSRRGRVDGMEPRGNAQVNAV	645
Qy	692	VLNDMFGYSTELRSCTEGSGEYTMESRYOPCLPSTOEVIK	735
Db	646	VLSEMFGYATSLRSNTGGRGTYTMFDHYAEVPSIAEELIK	689

Search completed: July 14, 2003, 18:27:01  
Job time : 21.6298 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:11:27 ; Search time 23.0933 Seconds  
(without alignments)  
2620.804 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730  
Sequence: 1 FCLQCTRWLRENGHPPST.....MIVKKEVSTRSASSQSSSR 2057

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1492	13.9	2548	4	US-09-172-422-1
2	1274	11.9	1939	4	US-09-310-187A-1
3	1232.5	11.5	1120	4	US-09-147-404-1
4	1065	9.9	1886	4	US-08-938-105-3
5	521	4.9	103	4	US-08-905-223-395
6	265.5	2.5	3248	1	US-08-353-700-1
7	265.5	2.5	3248	5	PCT-US95-16216-1
8	264	2.5	2482	1	US-08-328-254-6
9	255	2.4	1507	3	US-08-929-329-5
10	253.5	2.4	1162	4	US-08-728-323A-2
11	253.5	2.4	1162	4	US-09-298-568-2
12	244.5	2.3	1898	1	US-08-056-200-94
13	244.5	2.3	1898	2	US-08-800-644-94
14	218	2.0	717	4	US-08-910-925-1
15	211.5	2.0	1375	4	US-09-722-139-2
16	211.5	2.0	1375	4	US-09-721-832-2
17	211.5	2.0	1375	4	US-09-721-689-2
18	210.5	2.0	1360	4	US-08-393-569-2
19	208.5	1.9	2154	2	US-08-841-349-4
20	207	1.9	1279	4	US-09-724-517-2
21	207	1.9	1279	4	US-09-641-807A-2
22	206.5	1.9	1279	4	US-09-723-096-2
23	206.5	1.9	1200	3	US-08-840-006-5
24	205.5	1.9	10182	4	US-09-134-001C-3159
25	205.5	1.9	743	4	US-08-910-925-3
26	199.5	1.9	740	1	US-08-257-073-5
27	197.5	1.8	514	2	US-08-960-022-14

28	197	1.8	1354	3	US-08-685-871-2	Sequence 2, App11
29	195.5	1.8	703	4	US-08-910-925-4	Sequence 4, App11
30	192	1.8	619	1	US-08-465-746-2	Sequence 2, App11
31	192	1.8	619	1	US-08-214-164-2	Sequence 2, App11
32	192	1.8	619	2	US-08-467-852A-3	Sequence 3, App11
33	192	1.8	619	2	US-08-246-636-2	Sequence 2, App11
34	192	1.8	619	2	US-08-247-491A-3	Sequence 3, App11
35	192	1.8	619	2	US-08-319-795-2	Sequence 2, App11
36	192	1.8	619	2	US-08-468-985-2	Sequence 2, App11
37	192	1.8	619	3	US-08-312-949-2	Sequence 2, App11
38	192	1.8	648	1	US-08-072-070-2	Sequence 2, App11
39	192	1.8	648	1	US-08-469-434-2	Sequence 2, App11
40	192	1.8	648	1	US-08-214-222-2	Sequence 2, App11
41	192	1.8	648	2	US-08-467-852A-2	Sequence 2, App11
42	192	1.8	648	2	US-08-468-718-2	Sequence 2, App11
43	192	1.8	648	3	US-08-247-491A-2	Sequence 2, App11
44	192	1.8	648	3	US-08-446-201-3	Sequence 2, App11
45	192	1.8	695	1	US-08-127-499A-23	Sequence 23, App11

## ALIGNMENTS

RESULT 1  
US-09-172-422-1  
Sequence 1, Application US/09172422A  
Patent No. 6300485

### GENERAL INFORMATION:

APPLICANT: Adams, Arwen E.  
APPLICANT: Chiu, Choi Ying  
APPLICANT: Dahl, David  
APPLICANT: Gorman, Susan W.  
APPLICANT: Leng, Song  
APPLICANT: Sheffield, Val  
APPLICANT: Welch, Juliet  
TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED  
TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,  
FILE REFERENCE: 200130.442  
CURRENT APPLICATION NUMBER: US/09/172,422A  
CURRENT FILING DATE: 1998-10-14  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2548  
TYPE: PRT  
ORGANISM: Homo sapien

Query Match 13.9%; Score 1492; DB 4; Length 2548;  
Best Local Similarity 24.3%; Pred. No. 5.2e-107;  
Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;

QY	35	YGVFTYKQSTTHOKVTA--MHPTNEGVDDMASLTLEHGSIWYVNFQRYKKNQIWT	91
DB	118	YGSLSWLRVTRERRRMERGFLLPQPOKDFDLCSLPDLNKTLELNRDRFKHEKIYT	177
QY	92	YIGSLIASNPQPLAGLEPATMEQYSRRHGLPPHIFALANECYRLMGRHNOCLT	151
DB	178	YGSLLIYINPER-FLPIYNPYVMYDNDHOGKEPHIYAVADVAYHAMLQRKKNQICIV	236
QY	152	IKGSGAGKTEBTKLILKFLSVISQOSLELSLKETSCEVERAILESSPIMEAFNAKTYV	211
DB	237	IGSGSGSGTGQSTNPLHLHLTLSQGF-----ASGVEQIILGAGPVLERFNAKTH	289
QY	212	NNNSRFGKPVQVNTCQKGNIOGRIYDCILSSQNRVVRQNGERNYHIFVYLLAGLEHE	271
DB	290	NNNSRFGKPIQVNYQETGVGA-YVERKYLEKSRLLVQEHNEHRYHVFYLLAGASBD	348
QY	272	EEBEYLYSPREYHNLQ-----SGVEDEK-----TISDSE-----SPREYITAMDV	313
DB	349	ERSAPHLQPEBYHYLNOITTKKPLRQSWDDYCYDSEPDCTVGEGLRHRDFRLQULAMEW	408

QY 314 M0SKEVEVSHLAGILHAGNI-----EFTAGAOVSFKTLAGSABLLGLDPTQILD 369  
 Db 409 VGLPRTBROIFSLSLALHAGNI CYKKKTYRDSIDICNPEVLPVLSBELLEKEMLE 468  
 QY 370 ALTORSMLRGEILPPLANVOAVDSRDSIALMALYACCEBWTAKTINSRKINRKNEDK--- 426  
 Db 469 ALVTRTVTVGEKLIIPYKLAELAVTRNSMAKSLYALPDWIVFRINHALLNSKDEHNT 528  
 QY 427 ---SIGILDFGENFENVHFEQFNINYANEKLOEYFNKHIISLEOLEXREGLEWEDID 483  
 Db 529 KTLISGLVDLFGEDJENNNSFEQFCINFAERLQHFNOHIFYLEQEBRTESISHAND 588  
 QY 484 MIDNGECLDIEKK-LGLLALINESHPQATDSTLEKLSQHANNHFVYKPRVAVNPF 542  
 Db 589 YIDNTCCINLISKKPTGLHLHDESNFQATNTQTLIDKFKHQHEDNSYIEPPAVNEPAF 648  
 QY 543 GVNHYAGEVOYDVRGLERKRDTPRDDLLNLRESFPDYLD- 585  
 Db 649 IIKHYAGKYVYGVDFREKNTDHRPDIVALLSSKNAFISGMIGIDPVAVFRMALLRAF 708  
 QY 586 -----FEHVSNN-----NODTLKCG----- 601  
 Db 709 FRAMAFAREAGKKNIRKGTGHDTPALCALIKSMDSFLOHPVHORSLEILQRCKEEKS 768  
 QY 602 ----- 601  
 Db 769 ITRKNPRTPLSDQNMALNEKNQHDTPDIAMNGRTGIRQSRLSSGTSLLDKXGIPANST 828  
 QY 602 -----SKRRAP 607  
 Db 829 SSKLTERAHGILTRNNKPKSKPALPGLLEVNSLKLTRLTLQDRITKSLHLHKKKKPP 888  
 QY 608 TVSSQKVDLSHSLMATLSSSNPFYRCIKPNQKMPDQDAVNLQRLYSGLMETVRI 667  
 Db 889 SISAOQA-SLSTLMTLGOAEFYFKCIRSNABKPLKRSVDVLRQLRITMLTVOI 947  
 QY 668 KKAQVARRPPODYFRKYVLMENLALPEDVRGKCTSLQYDASNEWQKTKVFLRE 727  
 Db 948 RQGSYSKTSFODPVSHFVFLPRNIIPS--KENIQDPFKINLNDPNYQVGTWVFLKE 1005  
 QY 728 SLEQLEKREBEV-----SHANVITRAHVTGFLARKQYRK- 763  
 Db 1006 QERQHLQDILLHQVLRRIILQRMFVLLCRQHFHLRQASVLIQFMWNYLNOQVORDA 1065  
 QY 764 -----VLYCVIIOKRYAFILRRRFLHLKKAIVFOKORGLIAR----- 805  
 Db 1066 AYOXKAFWASAAALLQASWRAHLERQYILBPAALIVIOQKRDRYRRRHMAALICIQAR 1125  
 QY 806 --VYRQLLAEKREQEBEKKQEBEKKREBEREREREREAELRAQOEB-----ETRK 857  
 Db 1126 WKAYRE---SKRYQORKKIILLQSTCRGFARQFKALKEQRLRETKPEVGLVNIKGYG 1182  
 QY 858 QOELHLOKSO--KEALITRELEKQENKQVEITLREKEIEDQNKKE-----QOELSLRE 912  
 Db 1183 SLEIQSDPSEWEDCSFDRIRIKALIECKSVITESNRISRE--SSVDCIKESPNKQOEAQSOQ 1241  
 QY 913 ASLOKLOE-----RQOELRLREEBACRAQOELFELSLNFEIDECVANIIE---RSLSGG 963  
 Db 1242 SGVD-LQEDVLRERERSLLEDLHQKVGAKR--ESRRKRELOALFSLBLKVRSLGRI 1298  
 QY 964 S-----EFSSELAESACEE--KPNFNFSQ-----PYPEEB-----VDEGEADD 1001  
 Db 1299 SPSEDRMRSTELVPEGLQSPRGTPDESQSLLELTSYESQSKLESYISDBG-----D 1353  
 QY 1002 AFRDSDNPEBHGSDORTSGIRSDSSR-----EDYVMDTVPTPSADSTVLAPSQ 1057  
 Db 1354 LQPPSPKISSPFRDRNALASNSBTSNAHLKQDTMKEWVCSSES-----ITCKPQK 1409  
 QY 1058 DSGSLNSSSGESTYCMQONAGLPPDPGDYVDODYEDGATSGSSVTFNSYSQOMS 1117  
 Db 1410 DS---FISNLPFFIIPQODPLKTMISQDLSIQKMLLENBTJAEALTLIDINRETR-- 1464  
 QY 1118 PDYRCS-----VGTNNSGAYRPSSEGAQSSPDESEDEPDRPTDDEL----- 1161

Db 1465 -RYHCSGKQOIVPSINTESSNPVLKKLEXTANTEKEERQKLOQONKEKEMEQIROQTDIL 1523  
 QY 1162 -----SYRDSYVSCVTLIPYHSLFLYMKGGLNMSWRKRCVAKDEFTFLFRKQOE-- 1211  
 Db 1524 EKERRAKFTEKPRIGECILVAP-----SSYQSKORVERPSSLSLNTSNKGE 1570  
 QY 1212 -----ALKQGLMKKGGSGSTLSRBNWKRMFVLROSKLMYFENDSEBKLGTV----- 1260  
 Db 1571 LNVLSGLSKDAALAQKSSSAHLPPKD-----RVTYFFERKSPGSSSTYKELS 1621  
 QY 1261 -----EVRNAKEIIDN-----TTKEN-GIDI 1280  
 Db 1622 KTRMGTOQLNVAACKLSNNRISKREHFRPTQSYSHNSDDLREGNARPIEFTKDNMSIPTL 1681  
 QY 1281 IMADRTFHILASP-----EDASQW--FSLQSVHASTQEOIOMHDEQONPONAGTLDV 1334  
 Db 1682 VKKA--LNSKNPOLHKEDEPAMKPVKLAGEQETQRFSSV--DEQAKLHK--TWSQ 1734  
 QY 1335 GLIDSVK-----ASDPDRPNSFVITANRVLHCNADTEEMHWTLLQRSKGTRE-- 1388  
 Db 1735 GEITKLAARQKASDDDIRQR-----AKRFPKAGQGEKKTIRVKT 1777  
 QY 1389 -----GOEFT-----VNGWLHKEVKNSPKMSLKLKKRW 1417  
 Db 1778 TQSEVSPLEAGTDVIPAQFPDELAAYHTPPLSPBLPGSCRKEFKENKEPSP----- 1830  
 QY 1418 FVLTHNSLDYVSSSKNMLKGLTVLNSLCSVPPDEKIFKSTGYVNTVYRKHCYRLX 1477  
 Db 1831 -----KAKRRSVKJNSVALDSM-----HMQ----- 1851  
 QY 1478 TKLNEATSWSSVIONVTDTKAPIDTPQOLIQDI-KENCINSDVVEQIYKSN-PILRYT 1535  
 Db 1852 -----NDSVQ---IASVSDLAS-MDEFLKKYNDLNDSDSKDQTLVDVYFKKALKEFRGN 1903  
 QY 1536 HHPHSLPLPLPYGINLNLKQKGYTLODEAIKIFNSLOQLESMSD-PIPITGILQIT 1594  
 Db 1904 IFSFYSALLAMDG---KSIRYKDYALLFEQILEKTRMLEQRDLSGESPVAV-----W 1953  
 QY 1595 GHDLRPLDELYCOLIKON-----IVP-----HPSGNGNLYSMQILN- 1632  
 Db 1954 VNTFVPLDE-VMEKFTSDCTATVPKTERKRRKKETDLVEBHNGHIFKATQYSLPTV 2012  
 QY 1633 CLSCTFLPSRGLIKYLKFKLREQPGTEMEKALFTYESLAKT-KC-----R 1681  
 Db 2013 CEYCS-----SLWIMDRASVCKLCKYACHKKCKLKTATACSKKYDELSR 2059  
 QY 1682 EF-----VPSRDE-----IEALHROEM---ISTVCHGGSGCKITINSHTAGEVE 1726  
 Db 2060 QGVELSRLTSEDRVPLVVEKLIYVIMHGLYTEGIVKSGSTNKI-----K 2107  
 QY 1727 KLIRGLAME-DSRNNFALFEVNGHVDKALIESRTVADVLAKREKLAASEVGDLPWKFY- 1784  
 Db 2108 ELRQGLDTPAESVN---LDDVNIH-----VIASVFKM-----LKDLPNPLMT 2147  
 QY 1785 FPLY 1788  
 Db 2148 FELY 2151

RESULT 2  
 US-09-310-187A-1  
 ; Sequence 1, Application US/09310187A  
 ; Patent No. 6358751  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benichou, Gilles  
 ; APPLICANT: Fedoseyeva, Eugenia  
 ; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac  
 ; TITLE OF INVENTION: Graft Rejection  
 ; FILE REFERENCE: UCSF-090  
 ; CURRENT APPLICATION NUMBER: US/09/310,187A  
 ; CURRENT FILING DATE: 1999-05-12  
 ; NUMBER OF SEQ ID NOS: 3

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-310-187A-1

```

```

Query Match      11.9%; Score 1274; DB 4; Length 1939;
Best Local Similarity 24.8%; Pred. No. 3.7e-90;
Matches 452; Conservative 334; Mismatches 640; Indels 396; Gaps 59;

```

```

27 GIVRTDYGVGYFTYKQSTITTHOKTAMPTNEBEGDDNLSUTELHGGSIMNLTQRYKR 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GVIATEENKGTATVATEDVLOQ---NPKFDKQDMMLTFLPEAPVLFNLKERYVA 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 NQIMVYIGSLASVNPYOPFAGLYEPATMEQYSRRLHGLPPHIFALMECTRCMLKRRD 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 NMIVYISGLFCVAVNDEYKLP-VYNAEVAAYRGRKRSEPPIHFIISDNAYQYMLTDRE 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 NQCILIKGSGAGKTESTKLILKF--LSVISQOSLELSIKETSCEVERALIESPIMEA 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 NOSILITGSSGAKTNTKRVIOYFASIALIGRGKKNANNGKLEDOITIOANPALEA 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 FENAKTVNNSSRFGKTYOLNICOQNTQGGRIYDCILSQNRVVRQNPGERNTHIFYA 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 FENAKTVNRDSSRFEGKFIHFGATGLASADI-ETYLLEKSRVIFQLKAEERNYHIFYQ 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 LLAGLEHEREEFYI-STPENHYLNQSGCEDKTSIDQSPREVIYAMDWQFSKEEVR 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 ILSNKKPELIDMLLVTNNDPYDAFVSQ-GEVSVASIDSEELMANDADVDVGFTESEKA 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 EYSRLLAGILHGNIEF-----ITAGGAGVSFXTALGRSALLGLDPTQLTDLATQR 374
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 GYVYKLTGALMTHGNMFKOKOREBOABPDGTEDADK-----SAYLMGLNASADLLKGLCHP 403
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 SMFLNGEELTPLANVOAVDSRDSLAMALYACCFEYVIRKINSRLKGNEDFK-SIGLIDI 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 RYVNGREYVTKGQSVOQVYYSIGALAKAYEKMFMNMTKINATLETQPROQFISGLDI 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 PEPENEVNHPFOPFNIYANNEKLOEYFNHGISLBOLEYSGRELWVEDIDM-IDNECIDL 492
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 APEELFDFSPFOQLCINFTEKLOQFPNNHMFVLEBEEYKKEIETFTFDGMDLQACID 523
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 LLEKTLGLALINEESHFOATDSTLEKLSQH--ANNHFFVKRVAVN-----NFGVK 545
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 LLEKPMGINSIEEBECMFKATDMTKALYDNHLEKSN--FOKRNKIKGQEAHFSLI 581
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
546 HYAGEVOYDVGRGLEKRDTPRDDLLNLLRESRFDIYDLFEHVSRRNQDTLKC-GSRH 604
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 HYAGTVDNILGLLEKXKDPINETVALYQKSLKLMATLFSYATADTGDGSKSGK 641
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 RR---PTVSQKRVDSLSHLMATLSSNPFYVACIRPMQKMPDQDOAVLANQLRYSGM 661
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 KSSSPQTVALLHR-EMLNKLMTLRTTHPHFVACIIPNEKKAQVWVNDPLVHMQLECNV 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
662 LETVRIRKAGVAVRPFODFYKRYKYLMENTLALPE---IVRGKTSLSQLVDASNSQ 717
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 LBGIRLCRIGFPFRILYGFQRYRIL-NPVAIPBQGFHSRGTCKLSSLDIDHNOYK 759
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 LGKTVPLRESLEQKLEKREBEVSHAANVIRAHVGLFRAKOYRKVY---CVVILIQK 774
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 FGHTVKVFPAAGILGLLEEMRDERLSRIITRMQOAGQMLRFEFKIYVRDALVIQNN 819
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 YRAFL-----LRRFLHLK 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 IAFPMVQKMPWPKYFKI KPLKSAETEKAMTKKEFGRIKETLEKSEARKKELEKK 879
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 AAVFOK---QLRGQ-----IARVYRQLLAKRQOEKKOE--- 826
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 VSLQKRNLDQLOVQABQDNLANDAERCDLIKRI--QLEAVYKEMERLEBEEMNAE 937
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 ---EKKRREERERERREARELR-AQOE---TRKOE 860
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

938 LTAKKRKLDECSSEIKDIDLETLTAKEKEKATENTKVNKLEBAGLDEIITAKTKE 997
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
961 LEALQSQKEALBTLEKQKQENKQVEEILLEKEIEDLOQMEQO----- 906
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
998 KKALEAHQALDLDQVEEDKVNLSKSKVLEQOVDLESLBEKRYRMDLEBAKRL 1057
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
907 --ELSTEASLQ-----KLERRDQELRL 929
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1058 EGDLLKTOESIMDLNDKLOLEEKLKKEPDIQNSKIEDEQALALOLOKKALENQARI 1117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 EE-----BACRAQOFLESLNFD--EIDECYRNIERSLGGSEFSSSLASACEKNF 981
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1118 ELEEELAEARARARARVLEKLSDSLRELELSEERLEA-GGATSVQIEM---NKREA 1171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 NFSQPYE-EVYDEGEFADDDAF--KDSPPNSEHGS-----D 1016
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 EFQKRRLDEETLLOHEATTAALRKHADSVALEGEQDINDLQRYVKQLEKESPEKELD 1231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1017 QRTSGI-----RTSDSSE-----DPYNDTVFTSPSADTVLLAP 1054
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1232 DVTSMNEQIIRAKANLEKVSRTLEDQANEYVKLEBAQRSINDFTQRAKQTENGELAR 1291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1055 SVQDSGLHNS-SGSESTYCNPN-----AGDLPSFGQDIDYDQDDIED 1097
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1292 QLEKEALISQLTRKLSYTOQMEDLKQLEBEGKAKNALAALQSAHDCDLAEQYER 1351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1098 GAITSG---SVTFENSYGQSPDYRCSV--GTNNSGAVRFSBGAQSPDESEBPD 1152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1352 ETEAAEALQRYLSKANSVAQMRITYETDALORTLEBEAKKLAQLODA-EBAVEAVN 1410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 SRFDTDELISYR-----RDSVYSGVTLPEHSPFLYMGKLNNSWKRW--- 1195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1411 AKCSLEKTRKRLQNEIEDLVNDVVERSNAAAALDKQRF---DKILAEMKQYEBEQ 1466
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1196 -----CYLKQETFLMRSKQALKQWHLK-----GGSSSTL 1228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 SELESSQKARSLSLTELFKLNAAYESELHEETFGRENKNTQEBISDLTEQGBEGKXVH 1526
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 SRNWKQKMPVLR---QSKL-----MYFENSEBKKTGVEYRTAKIETIDNTKENGIDT 1281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1527 ELEKTRKQLEVEKLELOALBEAELSLEHESKILRAQLEFPQIAELIRKLAEDEME 1586
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1282 MADRTFHLIASEPE---DA-----SQWPSVLSQVHAS-TDOEIQEMDEQ--ANPQNAV 1330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1587 QAKRNHQVNVSLQSLDAETRSRNEVLRVKKMGEGDLENBEOUSHANRMAABAQOYK 1646
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1331 TL-----DVLGDISVCASDSPDRPSFVITIANRVLHCNADTBEMHMTTLQSKG 1383
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1647 SIQSLKQDQIOLDADAVANDDL-KENIAIVERNNLLQAEI---BEIRAAYEOTERSR- 1701
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1364 DTRVGEQEFYRWGMLHKEVNSPKMSLKLKRWFLVTHNSLDYKSSBKXALXIGTLVL 1443
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1702 -KLAEQELI-----ETSERVQLHSQNT-----SLNQKKNESDLOQSEVE 1744
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1444 NSLCGVNPPDEKIFKETGYMNVTVYGRGHYRLYTKLINEATRWSSVIONVDTAPRID 1503
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1745 EAVQCRNAEKA-----KCAITDAMMAEUKKEDDTSAHLER 1783
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1504 PTQOLIIDIKENCNSDVVEOI 1525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1784 MKKMMQGITKDLQHLDEAEQI 1805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-09-147-404-1
; Sequence 1, Application US/09147404
; Patent No. 6326163
; GENERAL INFORMATION:
; APPLICANT: FORSMANN, WOLF-GEORG
; APPLICANT: RAIDA, MANFRED
; APPLICANT: BRENNER, BERNHARD
; APPLICANT: NIER, VOLKER
; TITLE OF INVENTION: A METHOD FOR THE DIRECT DIAGNOSTIC DETECTION OF

```

TITLE OF INVENTION: GENETICALLY CAUSED PATHOGENIC POINT MUTATIONS  
FILE REFERENCE: 10496/P63211U0  
CURRENT APPLICATION NUMBER: US/09/147,404  
CURRENT FILING DATE: 1999-04-06  
NUMBER OF SEQ. ID NOS: 3  
SOFTWARE: PatentIn Ver. 2.1  
SEQ. ID NO. 1  
LENGTH: 1120  
TYPE: PRN  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (26)  
OTHER INFORMATION: "Xaa" represents Ala or Val  
NAME/KEY: MOD RES  
LOCATION: (59)  
OTHER INFORMATION: "Xaa" represents Val or Ile  
NAME/KEY: MOD RES  
LOCATION: (143)  
OTHER INFORMATION: "Xaa" represents Arg or Gln  
NAME/KEY: MOD RES  
LOCATION: (245)  
OTHER INFORMATION: "Xaa" represents Arg or Gln  
NAME/KEY: MOD RES  
LOCATION: (256)  
OTHER INFORMATION: "Xaa" represents Gly or Glu  
NAME/KEY: MOD RES  
LOCATION: (403)  
OTHER INFORMATION: "Xaa" represents Arg, Gln, Leu or Trp  
NAME/KEY: MOD RES  
LOCATION: (453)  
OTHER INFORMATION: "Xaa" represents Arg or Cys  
NAME/KEY: MOD RES  
LOCATION: (513)  
OTHER INFORMATION: "Xaa" represents Phe or Cys  
NAME/KEY: MOD RES  
LOCATION: (584)  
OTHER INFORMATION: "Xaa" represents Gly or Arg  
NAME/KEY: MOD RES  
LOCATION: (587)  
OTHER INFORMATION: "Xaa" represents Asp or Val  
NAME/KEY: MOD RES  
LOCATION: (602)  
OTHER INFORMATION: "Xaa" represents Asn or Ser  
NAME/KEY: MOD RES  
LOCATION: (606)  
OTHER INFORMATION: "Xaa" represents Val or Met  
NAME/KEY: MOD RES  
LOCATION: (615)  
OTHER INFORMATION: "Xaa" represents Lys or Asn  
NAME/KEY: MOD RES  
LOCATION: (716)  
OTHER INFORMATION: "Xaa" represents Gly or Arg  
NAME/KEY: MOD RES  
LOCATION: (719)  
OTHER INFORMATION: "Xaa" represents Arg or Trp  
NAME/KEY: MOD RES  
LOCATION: (723)  
OTHER INFORMATION: "Xaa" represents Arg or Cys  
NAME/KEY: MOD RES  
LOCATION: (731)  
OTHER INFORMATION: "Xaa" represents Pro or Leu  
NAME/KEY: MOD RES  
LOCATION: (736)  
OTHER INFORMATION: "Xaa" represents Ile or Met  
NAME/KEY: MOD RES  
LOCATION: (741)  
OTHER INFORMATION: "Xaa" represents Gly, Arg or Trp  
NAME/KEY: MOD RES  
LOCATION: (778)  
OTHER INFORMATION: "Xaa" represents Asp or Gly  
NAME/KEY: MOD RES  
LOCATION: (797)

OTHER INFORMATION: "Xaa" represents Ala or Thr  
NAME/KEY: MOD RES  
LOCATION: (870)  
OTHER INFORMATION: "Xaa" represents Arg or His  
NAME/KEY: MOD RES  
LOCATION: (908)  
OTHER INFORMATION: "Xaa" represents Leu or Val  
NAME/KEY: MOD RES  
LOCATION: (924)  
OTHER INFORMATION: "Xaa" represents Glu or Lys  
NAME/KEY: MOD RES  
LOCATION: (935)  
OTHER INFORMATION: "Xaa" represents Glu or Lys  
NAME/KEY: MOD RES  
LOCATION: (949)  
OTHER INFORMATION: "Xaa" represents Glu or Lys  
US-09-147-404-1  
Query Match 11.5%; Score 1232.5; DB 4; Length 1120;  
Best Local Similarity 30.4%; Pred. No. 2.7e-87;  
Matches 338; Conservative 216; Mismatches 374; Indels 183; Gaps 33;  
8 VMLRENGHF-PTVNSCAEGIVPRTDYGQVFTYKQSTIITQKTAHPTNEGVDDMA 66  
37 VFPPDDKQEFVAKIVSREGGKXTAETEGKTVTKEQVWQ---NPPKFDIEDMA 91  
67 SLTEHGGSIMYNLFPQYKRNQIWTYIGSILASVNPQPIAGLYEPATWEOYSRRHIGEL 126  
92 MLTFLEHRAVNLNDRYSMMITYSGLFCTVTPYKLP-VYTPEVAAAXGKRSBA 150  
127 PHIFAIANEYCRLMKRDNQIILKBSGAKTESTYLILKFLSVISQSLSLK 186  
151 PHIFISIDNAYQWMLTRENQSIILTESGAKTWTNTRVQYPAVALAIG-DRSKDQ 209  
187 T---SCVEBALLESSPIMEAFGNATVYNNNSRRGKFPQNLICQKNIGGRIVDCIS 243  
210 SPKGKLTEDQIIOANPALEAFGNATVYNNNSRRGKFPXIHFGATXKLASADI-ETYL 268  
244 SONRVVRQNGERNHIFFYALAGLEHEBEREFYL-STPENHYLNQSGVEDKTSIDQE 302  
269 EKSRIYFQAKARDHIFQILSNKKPELDMILLTNPNYDAFISQ-GETTVAISIDAE 327  
303 SFREVITAMDVQFSKEEVEVSRLLAGILHGNIEF-----ITGGAQVSFKTALG 354  
328 ELMATDNMFADVLGFTSEENSMYKLTGAIMFGNKKFKLKOREQAEPTGTEADK--- 383  
355 RAABELGDPQTDLTALQSRMFLRGEELTLPNLNQAVDSDSLAMALYACCFPMVIRK 414  
384 -SAYLMGLNSADLLKGLCHPXKVGNEEYVTGQNVQOVIATGALAKAVTERMFMMVTR 442  
415 INSRIKGNEDFK-SIGILDIPGFENFEVNHFEQFINYANEKLQEVFNHIFSLQLEYS 473  
443 INATLETQPKQYFPGVDIAGFEI.FDNPSPQLDINFNTEQLQGFNNHNVVLQOEAK 502  
474 REGLVWEDID-WIDNCECLDIKKLGLIALINBSHPPOATSTLLEKLSQH---AN 528  
503 KEGIEWTFIDKGMIDQACIDILEKPMGINSILEECMPFKADMTFKALFLFNHIGKSN 562  
529 NHFVYKPRANV---NGCVHGYAGEVOYDVGILEKRDTRDDLNLNRSRDFIYD 584  
563 ---FQKPRNIGKPRAPHSLSLHYAXIVXNIIIGWLOKNDPLKEXVXGLGYQSSIXLST 619  
585 LFEHVS-----RNNQDTLKCGSKRRRPTVSSQFVDSLSHSLMATLSSNPFVCIKPN 639  
620 LFRANYAGDAPLEKGGAKKSSSQ--TVSLNHR-ENLNKMLNTLBRSTHPRFVCIIPN 676  
640 MQKMPDQFQAVVNLQRLYSGLLETVAIRKAGYAVRPFQDFYKXYKVMRLNALPE--- 696  
677 ETKSPGVMDNPLVMHQKNGVLEGRICRKGFPRIILYXDQXQYXIL-NPAALXEGGF 735  
697 -DVRGKCTSLQLYPANSSEWQLGKTVPLRESLEQKLEKREBEVSNAAMVIRAHVLCF 755  
736 XDSRKAEBKTLSSLDIDHNQYKFGHTKVPFKAGLLGLLEMXERLSRIITRIQASRGV 795



QY 968 S---ELAESACEKPNF-----NFSPYPPEEVDDEGFEADD 1002  
DB 1122 KMRDLLEATLQHATPAALRKHADSVAEIGQIDNLQVVKOLEKEKSEFLELDDVT 1181  
QY 1003 FKSPNSSEHGHSDQRTSG-----IRTSDDSEF-----DPTMNTVVPPTSAD 1047  
DB 1182 -----SHMEQIIKAYANLEKVSRTLEDANEXRYKLEBAQSLNFTTQAKLQT 1231  
QY 1048 STVLLAPVODSGSL--HNSSGESTYCMPON-----AGDLPSPDGPDYD 1090  
DB 1232 ENGEARQLEKEKELIMQLRKLTSTYQOMEDLKQLEEGKAKNALAHLOQARHDCDL 1291  
QY 1091 DDDYEDGAIITSGS---SVTFMSYSGQSPDYRCV--GTNSSGAYRFSSGAGSSFE 1145  
DB 1292 LRQYEEEMEKAKELQVLSKANSVAQMRKYETDAIQRTLELEAKKLAQRLDA-E 1350  
QY 1146 DSEEDPDRFDJDELSTYR-----RDSVYSCVTLVYFHFYLYMKGLNWSK 1192  
DB 1351 EAVEAVNAKCSLEKTYHRLQNETIEDLMVYERSNMAAALDKKQRF---DKILAEMK 1406  
QY 1193 RRV-----CVLKDETFMFRSKOBALKQGMWLRK-----1221  
DB 1407 QXVEESQSELESQOKERASLTSLFKKAYEBSLEHLETFKXENKLOEISDLTBOQL 1466  
QY 1222 GGGSTLSRRNMKKRWFVL---QSKL---MYFENDSEKLGTYEVRTAKEIIDNTTK 1274  
DB 1467 EGKKNVHELEKIRKQLEVEKLEQSLAEBAASLEHEEGKILBAQLEFNQIKAEIRKLA 1526  
QY 1275 ENGIDIMADRTFHLLAESPE-----DASQMFVLSQVASTD 1312  
DB 1527 EKBEHEQAKRNHLRVVDSLOTSLDETRSRNALRYKKMGEDLMEIQLSQAARILAS 1586  
QY 1313 QEIQEMHDEQANQNAVGLTDLVGLIDVSCASDPDRPNSFVIITANRVLHCNADTEEMH 1372  
DB 1587 EAQKHLKNAQAHKADQLOLD---DAVRANDL-KENIAVERRTLLQAL--EELR 1638  
QY 1373 HMTLLORSKGDTRVSGEFTVRCWLHKEVNSPKMSLKLKRWFLVTLNLSLDYKXSE 1432  
DB 1639 AVEEOTERSR---KLAEQELI-----ETSERVQLLHSQNT---SLINQKKMD 1680  
QY 1433 KNALKLGTLVNLSCSVPPDEKIFKETGYMNTVVGKRGCHLYTKLNLNARMSVIG 1492  
DB 1681 ADLSQLOTEVEEAVQECRNAEEA-----KKAITDAMMAEBLK 1719  
QY 1493 NVTDTKAPIDTPTQOLQIDIKENCINSDVVEQI 1525  
DB 1720 KEDTSAHLERMKKNMEQTIKDLQHRUDEARQI 1752

RESULT 5  
US-08-905-223-395  
Sequence 395, Application US/08905223  
Patent No. 6223029  
GENERAL INFORMATION:  
APPLICANT: Edwards, Jean-Baptiste D.  
APPLICANT: Duclercq, Aymeric  
APPLICANT: Lacroix, Bruno  
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
NUMBER OF SEQUENCES: 503  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,223

FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 395:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
ORIGINAL SOURCE:  
ORGANISM: Homo Sapiens  
TISSUE TYPE: Brain  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: -72...-1  
IDENTIFICATION METHOD: Von Heijne matrix  
OTHER INFORMATION: score 3.6  
US-08-905-223-395

Query Match 4.9%; Score 521; DB 4; Length 103;  
Best local similarity 100.0%; Pred. No. 2.4e-33;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 643 MPDQPDQAVNLQNLRYSGMLFTVIRKAGYAVRPPDFFKRYKYLMLRNLALPEDVRGKC 702  
DB 1 MPDQPDQAVNLQNLRYSGMLFTVIRKAGYAVRPPDFFKRYKYLMLRNLALPEDVRGKC 60  
QY 703 TSLDLYDASNSEWOLGKTQVFLRESLQKLEKREBEVSHA 744  
DB 61 TSLDLYDASNSEWOLGKTQVFLRESLQKLEKREBEVSHA 102

RESULT 6  
US-08-353-700-1  
Sequence 1, Application US/08353700  
Patent No. 5599919  
GENERAL INFORMATION:  
APPLICANT: YEN, TIMOTHY J.  
APPLICANT: PATNER, JEROME B.  
TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCORE PROTEIN,  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
STREET: 1601 MARKET STREET, SUITE 720  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:



SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
US-08-353-700-1

Query Match 2.5%; Score 265.5; DB 1; Length 3248;  
Best Local Similarity 18.4%; Pred. No. 6,4e-11;  
Matches 422; Conservative 362; Mismatches 764; Indels 741; Gaps 110;

136 ECVRCMLKHNOCILIKSGSAGKTESTLILKFLSVISQSLSLSEKTSCEVERALL 195  
DB DCSQDLKKRHHIEQNDKLSTKESKALL-----SALERKKE----- 608  
196 ESSPIMEAFGNATVYNNNSRFGKFVQLNICQKNIQ--GRIVDCILSSQ-----NR 247  
DB -----YELKEKTLTSCMSKSENEKILLTOMSEKKNLQSKINLETCLTKQILKSHRYN 663  
248 VVRQNGE-----RVNY-----IFVALLAGLEHEEREERF-----YL 278  
DB RVTTLTMDRENLSVEIRNLNVLDSKSVETQKLAAMELOKAEFSDQKHQEIENMCL 723  
279 STE-----NYHLNOSGCVEDKTIISDOSEFRAVITAMVMQSKKEVNE 323  
DB KTSQLTQVDELEHLQLLSNEIMDKORCYD-LHAEYESLRDLKSKASLVTNEDHQR 782  
324 VSRLLAGLILHNIETTAGAGQVSEKTLGRSAELLGLDPTQLTALTORSWFLRGEET 383  
DB ---SLLA-----FDQPRAMHSPANTIGEGSM-----PSERSECRLEADSPKASAI 827  
384 LTPV--NVQOAVDSRSLAMALVACCEVWIKKINSRINKNEDPKSIGLIDIFGFENFEV 441  
DB LQRRVDSLEPSLSEKQKMSDLOKCEELV-----QIKGEIE-----EN--L 867  
442 NHEQONINANKLOEYFKKIHPSLEOLEYREGVWEDIDWDINGECIDLIEKLGLL 501  
DB MKAEQNHQSFVATSQ-----RISKLOEDTSAHQNIV-----AETLSALENEKEEL 913  
502 ALINESHPQATDSTLLEKLSQHANNHFYKPRVAVNVNFGVHAYAGEVQVVRGILEK 561  
DB QLINDKVEIEQA-----LRESRDFIYL-----FEHVSNNODTLKSGSKRRPTV 609  
914 SNHLBDSLKEQLLSLETSLSEKKESSIIISLNKRIBELTQENG--TLK-----EI 982  
610 SSQPKVDSLHSLMATISSNPPFVRCIKENMQMPDFOQAVVLNQLRISGMLETVRIRK 669  
DB NSLINOBNKMLIKSSEFPAN--YIDEREKSIISLSOYIQOEKLI-----LIQ--RCEB 1031  
670 AGAVNRPDQFYKYVLMRLNALPDEVGKCTSLLOLYDASNSEMOGKTAVFLRESI 729  
DB 1032 TGNH-----YEDLSQTKAAQEKNSKLECLNCTSLCE-----NRKNEL-----EQL 1074  
730 EQLERKREBEVSHAAMVIRAHVIGFLARKQYKRVLCVVIQKNYRAPL----- 779  
DB 1075 KEAFALNHQFLKLA-----FAERNQNLMLLETVOQALRSEMTDNQNNSSK 1124  
780 ---LRRFLMLKKAALVFOQLAGQIARRVYRQULAEKVEBEKKKREBEER 835  
DB 1125 AGGLKQEIIMTLKEQNMQKEVND-----LIQENQMLKVMKTKIECNLTSEPT 1174  
836 ---REBERRRREAEALAOOEETRKQO-----ELBALQSKQKAEVLTRELKQK 882  
DB 1175 RNSVKERESERNQCNFPQNDLVEKLSLDSYNAQLVQLEAMLRN--KELKL--QESERKE 1232  
883 NKQVE--EILRLKEIEBIDLRMKEQO-----ELSLTEASLQRLQER 921

DB 1233 CLQHEIQTIRGDLTETSNLODMQSOEISGLKQCEIDAEKYLISGRHETLSQNDNHLQCS 1292  
922 RQGEIIRLE--BEACRAQ-----EFLSELNPEIDCVY-----NI 956  
DB 1293 LQTTMKLNELEKICEIILDAEKYELVTEILN--DSRSECTIATRKMAEYVKLLNEKYLND 1351  
957 ERSLSGSEFSSELASACEERKN-----FNFSQPY-----PIREVDGEFPAADDA 1002  
DB 1352 DSGLLHG--ELVEDIPGEGEGEQPNQHPVSLARLDESNSYBHLTJSDKEVQNHFLQLQEK 1410  
1003 FKDSFNP-----SEHSHDQRTSGIRTSDDSEEDPYRNMOT----- 1038  
DB 1411 FLSQSEHKILNDHQCMSSKSKSELQTYVDSIKANLVLSTLRRNFGDLVKEMQLGLEE 1470  
1039 -VPTSPASDSTVLLAPSVQDSGLHNSSGSESTY--CMPQNAQD---LPSPDGYVDQ 1092  
DB 1471 GLVPS-----LSSSCVPDSSSL--SGLGDSFYRALBQTDGMSLISLNEGAVSANO 1520  
1093 DDYED-----GATSGSSVTFPSN-----SYGSQMSPDY 1120  
DB 1521 CSVDVFPSSLOTYVDSLKAENLVSTNLRRNQGLVKEMQGLREGVPLSSSCVPD- 1579  
1121 RCSVGTNNSGAYR-----PSSGAQSSPEDS-----BEDFSRFTD- 1158  
DB 1580 SSSLSSIGDSSFYRALLBQTDGMSLISLNEGAVSANOCSVDVFPSSLOEENILTRKETPS 1639  
1159 -----DELS-----YRRDSVVSCTPLPYHPSFLMKGLGM--SWKRRKCVLKDPTLMP 1206  
DB 1640 APAKVEELSESLCEVYRQ-----SLEKLEEKMSQGIKMKKEIQBLEQLLSSB----- 1687  
1207 RSKQEBALQGMHLHKKGSGSSTLSRRNMKRWFLVQSKLMPENDSEKL-----KGVAV 1262  
DB 1688 ROEDLCRKQVYS-----NQW-----QQLUTSVTLMESSLAEKKQTBEL 1730  
1263 RTAKETIDNTKENGIDIMADRTHLIAESPEDASQMFV--LSQVHASTDQEIQENH 1319  
DB 1731 SLELEVA--RLQLOGLD--LSSRS--LIGIDPEDIAQGNESCDISKEHTSETTERTPRG 1784  
1320 D--EQANPQNAV--TLDDGLIDSV-----CAAD--SPDR----- 1348  
DB 1785 DVHQICDQAQODNLNLEKITEGTGAVKPTGCSGQSPDTNYBPFGEKDTQGSSECSB 1844  
1349 -----PNSFVIL-----TAN-----RVLHNCADTFEEMHMTLLQSRK 1382  
DB 1845 LSFSGBNALVPRDFLQNGEDINLQRYKEISNEMLRLHVLBDRKAY--SLINEMK 1901  
1408 MSS-----LKLKKRW-----FVLTHNSLIDYKSSFKNALKGLTVLANS 1446  
DB 1962 LNSDLEMADKSRREDIGDGNVAKVDSWKEKRLDVENEISRLR--SEKASIEHEALYLEAD 2020  
1447 CSVPRDEKIFKQGTGMVNTVVGKHCYRLVLYKLNEATRMSSVQVONTAPIDTPQ 1506  
DB 2021 LEAV--QTEKLCLEKQEN--KQVYVCELEELSVTSEKNQRLGELDTNSKTTALD 2074  
1507 QLIODICE-----NCLNS--DVVEQIYKRNILRYTHNPLHSPLLPLRYGDIINL 1555  
DB 2075 QLEKMKKEKTOLESHQSECHICQVABEAVEK-----TELLQTLSSDVS--EL 2122  
1556 LKQKGYTTLQDBALIKFNSLOLESMSDPIPIQIGILTQGHDLRPLRDBLYQQLIKQTK 1615  
DB 2123 LKDX--THLQEK-----LOSLEKDSQALSLTKCELB--NOIAQUNKEKL--LYESES 2170  
1616 VHPGSVGNLYSMOILTCISCFELPSRGLIKLYKHLKRIRQFPQTEH-----EKY 1667  
DB 2171 LQARLSSESYKELNYSKALBAALVKEGFAIRLSTQSEVYNQLRGIRIKLRIRADEK 2230  
1668 ALFTYESLKTQCRFPVPSRDEIALIHRQEWTSFVYCHGGGSCKITINSHTAGEV--- 1724

Db 2231 QLIHAEKLERE-RENDLSLKQYENIEREQMSEB-----NOELVILDAENSKAEVELT 2283  
Qy 1725 ---VEKLRIGLAMEDSDNMFALFEYNGHVDKAIERS---TVADVLAKEFKLATSSEVG 1777  
Db 2284 KTOIEBARKLKTPE-LDVLTLSEKENLTKOIOEKOGOLSELDKLSSKSLLEKEQA 2242  
Qy 1778 DLFMKFYFKLYCPLDIDNVKDSVEFAFMFEQAHEAVIGHHAPAEENQVIALRLQVY 1837  
Db 2343 EIQIK-----EESKTAEMQONQKKEINAEVAA-----L 2371  
Qy 1838 QGYTLAAAI-----PLEEYV-----SLQRLKARISSTKTFPCCELEKRRSPLEGLT 1888  
Db 2372 CGSOETIKATQESLDPRIEEHQURNSIEKLRARL-----EADKKQLCVLO-Q 2420  
Qy 1889 RRSFRITSVVRQVEEQMLDMWIKEEVSARASIIDKMRKFOQMNQOQAMAKYMALIKE 1948  
Db 2421 KESEHNADLKGVEN-----LERLEIART-----NOEHALE-----AEN 2457  
Qy 1949 WPGYGTLLFDVECKEGFPQELMLGVADAVSVYKGGEPLEVOYEHLISFGAPLANT 2008  
Db 2458 SKGEVEVL---KAKIEGMTQS-LGELDVVTIRSEKENLTNELQKEQERISELEIINSS 2513  
Qy 2009 YKIVUDERE 2017  
Db 2514 FENIQEKE 2522

## RESULT 7

PCT-US95-16216-1  
Sequence 1, Application PC/TUS9516216  
GENERAL INFORMATION:  
APPLICANT: Yen, Timothy J.  
APPLICANT: Ratner, Jerome B.  
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
TITLE OF INVENTION: Expressed Kinocochre Protein, and Methods of Use  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Releasee #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16216  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,700  
FILING DATE: 09-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match 2.54; Score 265.5; DB 5; Length 3248;

Best Local Similarity 18.4%; Pred. No. 6,4e-11;  
Matches 442; Conservative 362; Mismatches 764; Indels 741; Gaps 110;

Qy 136 EGYRCLMKRNDQCLLIKESGAGTSEFKLLKFLSVISQOGLSLKERTSCVERAIL 195  
Db 568 DCSQDLKREHTEIQLNDLSTKTESKAL-----SALKKE----- 608  
Qy 196 ESSPIMEAFGNAKTYANNNSRFGKLVOLNIOCKNIOG--GRIVCILISQ-----NR 247  
Db 609 -----YEBLKEKRTLPSCWKSSENEKLLTQMESEKENLOKINHLETCLKTQOIKSHEVNE 663  
Qy 248 VYRONPGE-----RNYH-----IFALLAGLEHERREEF-----YL 278  
Db 664 RYRTLEMDRENLSVIRNLHNLVLDKSVETQKLAHYELQKAEFSQKQKEIENMCL 723  
Qy 279 STPE-----NHYLNOGCGVEDKTIISDOSPFEVITAMDVMQFSKEVNE 323  
Db 724 KTSQTLQGVEDLEHKLQOLISNIMDRCTYD-LHAETESLRLDKSDASLVNEDHOR 782  
Qy 324 VSRLLAGLILHGNIEFTAGAOVSFKTLAGRSABELGLDPTQTLALTORSMPFRGEBI 383  
Db 783 ---SLLA-----FDQPAWHHSFANILIGQSM-----PSEBECRLADQSPKNSAI 827  
Qy 384 LTRPL--NVOQAVDSRDSLAMALYACCFEYVYIKINSRIKGNEDFFSIGLIDIFGEPNEV 441  
Db 828 LQNRVDSLEFSLESOKMNSDLQKOCSEELY-----QIKGEIE-----EN--L 867  
Qy 442 NHPEOFNINYANEKLOEYFNKHIIFSLQLEYSREGVIMEDIMNGEGLDLEKKGLL 501  
Db 868 MRAEQNHQSFAVETSD-----RISKLOEPTSAHONV-----AETLSLENKEKL 913  
Qy 502 ALINESHPPOATDSTLLEKLSHONHNFYKPVAVANNFGVAYAGEVOYDVGILIEK 561  
Db 914 QLNLDKVERTEQA-----EIQE-----LKK 932  
Qy 562 NDTFRDILNL-----LRESRFDITYD-----FEVSRNNODTLKCSKRRP 609  
Db 933 SNHLEDSLEKELQILSETLSLEKKEMSSITSNKREIEELTQENG--TLK-----BI 982  
Qy 610 SSQFKVDSLSMATLSSNPFVRCIKPMOKMPOFOAVNLQRLYSGLMETVRIK 669  
Db 983 NASLNOEKMLIQKSEFPAN--YIDEREKISIELSDQYKQKELI-----LLO--RCE 1031  
Qy 670 AGYAVRPFQDFYKRYKVLNMLALPBDYRGKTSLLQLYDASNSQWQKTKVFLRESL 729  
Db 1032 TQNA-----YEDLSQYKAAQENKSLTECLINCTSLCE-----NKNEL-----EGL 1074  
Qy 730 EOKLEKREBEVSNAAMVIRAHVGLPLARKQYRKVLVCVVIQKRYAFL----- 779  
Db 1075 KEAFKKEHQEFPLTKLA-----FAEERQNLMLLETVOQALRSMTDNONNSKSE 1124  
Qy 780 ---LRRRLHLKKAIVPQKOLRGQIARVYNQALLAEKREBEKKOEBEKKREBE 835  
Db 1125 AGGLQOETWTLKEBQNKQKEVND-----LLOEBQMLKMYKTHQCONISEPT 1174  
Qy 836 ---REREREREAELRAQOEBETRKOQ-----BLEALQSKQKAEFLTRELEKQ 882  
Db 1175 RNSVVERESERNQCFKPMQMDLEVKEIGSDYNAQLVQLEAMLRN--KEKL--QESKEKE 1232  
Qy 883 NKQVE-EILRLKEKIEDLQRMKEQ-----ELSITEASLOKLOER 921  
Db 1233 CLQHEIOTRGDLETSNLQDMQSOEISGLKCEIDAEBKYSIGPHELSTQNDNAHLOS 1292  
Qy 922 RQGLRRLLE--BEACRAQO-----EFLSILNPEIDECV-----NI 956  
Db 1293 LQTMKLANLEKTEICELIAEKYELVTELN--DSRSECTIATRKAAEYKGLNEXKILND 1351  
Qy 957 ERSLSGSEFFSSELASACEEKP-----FFFSQPY-----PSEVDEGEFADDA 1002  
Db 1352 DSGLIHG--ELVEDIDGEGEGEPQNHQVSLARLDESNSEYHLLTSDKEVQNHFAELQEK 1410  
Qy 1003 FKDSFNP-----SEHGHSQDQTSIGRTSDSSEEDPYANDT----- 1038

```

Db      1411 PLSLOSEHKLHDQHCQSSKSELOQYVDSLKAENLVLTNIRNFQGLVEMQGLSE 1470
Qy      1039 -VPTSPASDSTVLLAPSVODSGSLNHSSESTY--CMEQONAGD--LPSPDGDYDQ 1092
Db      1471 GLVPS-----LSSSCVPDSSSL--SLIGDSSFRLALIEQDMSLSNLEGAVSANO 1520
Qy      1093 DYED-----GAITSGSSVTFSN-----SYGQSPY 1120
Db      1521 CSDVEFCSSLQTYVDSLKAENLVLTNIRNFQGLVEMQGLSEGLVPSLSSCVPD- 1579
Qy      1121 RGSVGTNNSGAVR-----FSSGQSSFEEDS-----BEDPFSRDTD- 1158
Db      1580 SSSLSLSDSSFFRLALIEQDMSLSNLEGAVSANOCSVDVFCSSLOEENLTREKPS 1639
Qy      1159 -----DELS-----YRDSVYSCVTLPYHPSFLYKGGIMN-SMKRRMCVLKDETFLMF 1206
Db      1640 APRAKVEELSECEVTRQ-----SLEKLEKMEGQGINKEIOLEOLLSSE----- 1687
Qy      1207 RSKQKALKQGMWLHKKGSSSTLSRRMKRPFVLRQSKLHYFENDSEKL---KGTVEV 1262
Db      1688 RQELDCLRKQYLSSE-----NEQM-----OQKLVSVTLMEWSKLAARKQTEOL 1730
Qy      1263 RRAKEIIDYTKENGIDIMADTFHLIASPEDASQWRSV--LSQVHASTDQEQENH 1319
Db      1731 SLELEVA--RLQLOGLD--LSSRS--LLGIDTDAIQGRUESCDISKHTSETTERTPKH 1784
Qy      1320 D-EQANPQNAV--TLDVDGLIDSV-----CASD-SPDR----- 1348
Db      1785 DVHQICDKQADQDLNDIEKITETGAVKPRGEGSGSPDTNYPERGEDTKQSSSECTSE 1844
Qy      1349 -----PNSFVIT-----TAN--RVLMCNADTPEEMHMITLLQRSK 1382
Db      1845 LSPSGNALVPMDFLNOEDIHNLQVRKETSNEMLRLVIEDROKVE--SLINEMK 1901
Qy      1383 G-DTRREGGF-----YRGMLHKE-----VQNSK 1407
Db      1902 ELDSKHLDEVOVMTKIEACIELEKIVGELKENSPLSELEVPSCDHOELLQORVETSEG 1961
Qy      1408 MSS-----LTKKKRM--FVLTHNSLDYKSSSEKNALKGTLYVNSL 1446
Db      1962 LNSDLEMAHDAKSRREDIGNVAKVNDWKRFLDVENELSRIT-SEKASIEHVALYLEAD 2020
Qy      1447 CSVPPDEKIFKETGYMNTVYGRKCYRLYTKLNEATRMSSVIONVDTKAPIDTPQ 1506
Db      2021 LEVY-QTEKLCLEKQEN-----KQVYVCLSEELSVTSERQNLGELDTMSKTTALD 2074
Qy      1507 QLIQDIKE-----NCLNS-DVVEQIYKRNPIRLRYTHPLHSPPLPLPYGDINLNL 1555
Db      2075 QLSERKKEKTOBLESHQSECLHCIOVAEAEVKEK-----TELLQTLSSDVS-EL 2122
Qy      1556 LMDKQYTTQDEAIKIFNSIQLESMSDPIPIQIGLQTHDLRLADELYQOLIQNTK 1615
Db      2123 LMDK--THLOEK-----LOSLEKQSAISLTKCELE--NOQAQNKKEKL-LVKESES 2170
Qy      1616 VPHPSGVNLYSWQILTCLSTFLPSRGLIKYLKFLKRIEOPFSTEM-----EKY 1667
Db      2171 LQARLESSEYBKLAVSKALEALAVKGEFALRLSRTOEYHQGRGIEKLARVLEADEKK 2230
Qy      1668 ALFTYSLKTKREVPSRDELEALIHROEMSTYVCHGSGCKTINSHTTAGV--- 1724
Db      2231 QLIHIAKLEKERE--RENDSLDKYENLERLOMSEB-----NOELVILAENSKAVEVTL 2283
Qy      1725 ---VEKLIRGLAMEDSRMFPALFEVNGVDKAIESR---TVAVDYLAKPEKLAASENG 1777
Db      2284 KIOIEEMASLKFPE-LDLVTLSEKENLTKQIQEQGCLSELDKLSSFKSLLEKEQA 2342
Qy      1778 DLPMKVFYKLYCFLDIDNVPKQSVBPAFMPEQAEHAVIIGHHAPAPENLQVLAALRQVYL 1837
Db      2343 BIKIK-----EESKTAVENTMQLQNTKELNEVANA-----L 2371
Qy      1838 QGYTYLHAAL-----PLEBEVY---SLQRLKARISOSTKTPPCERLEKARTSPLEGLT 1888
Db      2372 CGQOETMKATBOSLDPPIEEBHQLRNSIBKRLARL-----BADEKKQCLVIO-QT 2420

```

```

Qy      1889 RRSFRTGSVROKVEEQWLDMMIKEVVSARASIIDKWRKEQGNVQEQAMKYNALIKE 1948
Db      2421 KESEHNAHLKRVEN-----LERELEIART-----NQERDALE-----AEN 2457
Qy      1949 WPGYSTLEFDVECKEGFPQBELMCLVSADAVSVYRGGRPLEVFOYTHILSFGAPLANT 2008
Db      2458 SKGEVETL---KAKIEGHTQSL-RGLBLDVVTIRSEKENLTJELQKQERISELBIINS 2513
Qy      2009 YKIVDERE 2017
Db      2514 FENILQEK 2522

RESULT 8
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match      2.5%; Score 264; DB 1; Length 2482;
Best Local Simlarity 18.4%; Pred. No. 5,4e-11;
Matches 401; Conservative 339; Mismatches 729; Indels 710; Gaps 102;

```

```

Qy      135 NECYCLMKRHNQICILKNG-----ESGAGKTESFKLILFLSVISQOSLESLKETS 189
Db      30 NERVATLEMDRENLSVEIRNLHNVLDSSVEVETQKLAMEL-----QOKAEFSQOKHKE 85
Qy      190 VERALIESPPIWEAFGNKATVYNNNSRFGKRVOLNICQKGNIOGRIYDCLSSONRVV 249
Db      86 IE-----NMCLKTSQLGQYED----- 102
Qy      250 RQNPGRNTHIFYALLAGLEHEREREFLSTPENRHYLNQSGCVAEKTIISQDESFRVIT 309
Db      103 -----LEHK-----LOLLSN-EIMDKORCYOD-LHAEYBSLRDLK 136

```

QY 310 AMDVQPSKEEVEBSVRLAGLILHGLNIEFTAGQVSPFKALGSAEHLGLDPTQD 369  
137 SKDASLVTNEDHQ--SLLA-----PDQPAHNSFANTIGEGSM-----PESRSE 181  
QY 370 ALQPSHFLGEBELTFL--NQOAVDSRSLMALYACFEVYIKINSRIGNEPDFS 427  
182 CRLEADQSPKSAILOVRVDSLESLESOKQMSDLOKOCBELV-----QIGGEIE-- 232  
QY 428 IGLIDIFGFENFEVNHFEQFNINYANEKLOEFNKHIFSELOEYSNEGLWEDIMID 487  
233 -----EN--LMKAEQWHSQFVAETSO-----RISKLOEOTSANQV----- 267  
QY 488 GEGLDLEKKLGLLALINEESHPPQATSTLEKHSQHANNHFYKPRVAVNPGVKHY 547  
268 AETLSALENKEKELOLNDKVETEQA----- 233  
QY 548 AGEVQDVQVIGLEKNRDTFRDILLN-----LRESRDPITYD-----FEHVSRRNQ 555  
294 --BIOE-----LKKSNHLDBSLKELOLSETLSEKEMSSIIISNKRIBELTOENG- 345  
QY 596 DTLKCSKRRPVTYSQFVDSLHSLMATLSSSNPFVRCIKPMQMPQDFQAVVLAQ 655  
346 -TLK-----EINASLNEKKNLLOKSSFPAN--YIDEREKISIELSDQYQOEKLI-- 392  
QY 656 LRYSGMLETVRIRKAGYAVRRPQDFYKRYKVLMRMLALPEDRGKCTSLQYDASNSE 715  
393 -----LLQ--RCSETGNA-----YEDLSQKYKAAQENKSKLECLINETSICE-----NKR 436  
QY 716 WQJGKTVFLRESLOKLEKREBEVSHAMVIRAHVGLFLARKQYRKVLVCVITIQKY 775  
437 NEI-----EOLKEFAKEHOFITKLA-----FAEERNOMLLELETVOAL 478  
QY 776 RAFL-----LRRFLHLKKAALVFOKOLRGQARVRVQQLAEKREOEKK 822  
479 RSEMTDNQNNKSABAGLKOELMTLKEONKMOKEVND-----LLOEBOQLKVM 528  
QY 823 QEEBEKKREBE-----REREREAREALRAQOEETRKOQ-----ELEALQSQ 868  
529 KTHGECONLESEPIRNSVKERESERNQCNFKPQMDLVEKISLDSYNAQVQJEAFLRN- 587  
QY 869 KEBELTRELKOKENKQVE--ELRLKEIEDLOPMKEQ-----E 907  
588 KEIKL--QESKEKECLOHELOTRIGDLETSNLODMOSQELSGKDCIDEAEKYSISPEH 646  
QY 908 LSLTEASLOKLOERRDOELRLE--BEACRAAO-----EPLSELNPEIDECVR----- 954  
647 LSTSONDNALQCSLOTTMKNLEKICEILOAEKELVTEIN--DSRSCITATRKMAE 705  
QY 955 -----NIESLSGSESEFSELAEACEKN-----FNFSQPY-----P 988  
706 EVGKLINVEKILNDSCGLHG--ELVEDIPGEGFQEQNEQHPVSLAPLDESNGEHLTIS 764  
QY 989 EEEVDGEFADDAFDKSPNP-----SEHGHSDQRTSGIRTSDDSEEDRYMNDT----- 1038  
765 DKEVQNHFAELOKFLSLOSEHKILHDQOMSSKSELOQTYVDSIKAEMLVISTULRNF 824  
QY 1039 -----VPTSPASDSTVLAPSVODSSGLNSSGSESTY--CMPONAGD- 1080  
825 QGGLVEMQGLGEGGLVPS-----LSSCVPDSSSL--SSLGDSFRLALLEOTGDMA 874  
QY 1081 --LPSPDGDYDQDDVEDGALITSGSVTFPSNYSYGSSOWSPDYRCSVGTYNSSGAYRPS- 1137  
875 SLTSLNEGAVSANQCSVDEVFGSSLOEENLT-----RKETPSAPAKVEELES 923  
QY 1138 --EAGSSFEDESSEDFPSRFTDELSTYRDSVYSCVTLPYFHSFLYMKGLNMSMKRMC 1196  
924 CEVYRQSLKLEKEMESQ-----GIM--TKKEI 949  
QY 1197 VLDEFTFLWRSQOALKQGLHKKGGGSSTLSSRMKMKRWFLRQSKLMTFENDSEBK 1256  
950 QELBQLLSERQSLDCKRQYLS-----NEQW-----OQKLTSTVLEHESKL 992  
QY 1257 ----KGTVEVTRAKEIINDTTKENGDIIMADRTFHLIAESPDAQWFSV--LSQVHA 1309

DB 993 AAEKQTEQLSLELEVA--RLQLOGLD--LSSRS--LLGIDTEADAIQGNESCDISKERT 1046  
QY 1310 STDQELQEHND--EQANPQNAV--TLDVGLIDSV-----CASD--SPDR----- 1348  
DB 1047 SETTERPRGHVQICDDKAQODLNDLEKILTEGALKPTEGCSGQSPDNYEPPEDEK 1106  
QY 1349 -----PNSFVIL-----TAN--RVLHCNADTPEEMH 1372  
DB 1107 TQGSSECSIELSPGPNALVPMDPLNGQEDIHNLQVRVETSENENRLHLHVEDDRKYE 1166  
QY 1373 HWITLQSKG--DTRVEGEF-----IVRGWLHKE----- 1401  
DB 1167 ---SLINKEKEIDSKLHLOEVQMLTKIEACIELEKIVGELKKEKNSDLSKLETFSCDHQ 1223  
QY 1402 -----VNSPKMS-----LKLKRR--FVLHNSLDYKSSBKAL 1436  
DB 1224 LLORVETSEGLNSDLEMTADKSSREDIGDNVAKVDSMKERFLDVENEISLRIR--SEKAST 1282  
QY 1437 KLGLVLNLSGVVPDEKIFKETGYMNVTVYGRKCYRLYKTLNEATRMSSVIONVTD 1496  
DB 1283 EHEBALYLEADLEV--QTEKLCLEKNEN-----KQKIVLCLEELSVTSENOJRGELD 1336  
QY 1497 TPAPIDPTQOLIDIKE-----NCLNS--DVEQIYKRNDILRYTHHPLHSLIP 1545  
DB 1337 TWSKTTALDQSEKKEKTEQLESHQSECLHICIOVAEAEVKEK-----TELLQ 1385  
QY 1546 LRYGIDINLNDKQYTLLODAIKIFNSLOQJESMSDPIPIQOGLQTHGLRLPLRDL 1605  
DB 1386 TSSDVS--ELKDK--THLOEK-----LOSLEDSQALSTTKELE--NOIAOLNKEK 1433  
QY 1606 YCOLIKQTKVPHPSVGNLYSMQILCTSCFLPSRGILKYLKRLKRIREQFGTEW- 1664  
DB 1434 EL--LYKESSELQARISESDYEKLANSKALBALVEKGERALRLSSTQEBVHOLRRGIEL 1492  
QY 1665 -----EKYALFTYESLKTKCREFPVBSRDEIALHROENTSTVYCHGGSCKITINS 1717  
DB 1493 RVRIBADEKQHLIAEKLEKERE--RENDSLKQVEMLERELQWSEE-----NOELVILDA 1545  
QY 1718 HTTAGBV-----VAKLRGLAMEDSRMFLFEVNGVDAKLESR--TVADVLKFE 1767  
DB 1546 ENSKAEVETLQOIEEMASLKVFE--LDLVYTRSEKENTQIOERQOGLSLDKLSSF 1604  
QY 1768 EKLAAATSEVGDLPMKFYFKLFCFLDTDVVPKDSVFAPFQDAHVAHGHHPABEENQ 1827  
DB 1605 KSLLEKEKAEIQIK-----BESKRAVENLQNLKELNBA 1639  
QY 1828 VLAALRLOYLOGDYTLHAI-----PLEEVY--SLQRLKARISOSTKFTTPEERLEK 1878  
DB 1640 VAA-----LCGDQOIMKATEQSLDPPIEBEHQLNRSIEKLRARL-----EADBK 1683  
QY 1879 RRTSFLGTLRSPFTGSVVRKXVEBQMLDMWIMEVSSAASIIDMKRKQGNQOBA 1938  
DB 1684 KQLCVIQA--QLKSEHHAIDLKRVEN-----LERELIAT-----NOEBA 1723  
QY 1939 MAKYMALIKEMPGYSTLFDVECKEGGFQOELMLGVSDAVSVYRGEGRPLEVQYEH 1998  
DB 1724 ALE-----AENSGEVETL--KAKIEGNTQSL--RELELDVYTRIBEKENLTLNELOKEQDR 1775  
QY 1999 LSFQAPLANTYKIVDERE 2017  
DB 1776 ISELINSSFENILOEKE 1794

RESULT 9  
US-08-929-329-5  
; Sequence 5, Application US/08929329  
; Patent No. 6120770  
; GENERAL INFORMATION:  
; APPLICANT: Adams, John H  
; APPLICANT: Dalton, John P  
; APPLICANT: Kappe, Stefan  
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing

TITLE OF INVENTION: Vaccine Compositions  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Barnes & Thornburg  
STREET: 11 S Meridian  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/929,329  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Breen, John P  
REGISTRATION NUMBER: 38, 833  
REFERENCE/DOCKET NUMBER: 835910-28685  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 231-7745  
TELEFAX: (317) 231-7743  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1507 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium yoelii  
US-08-929-329-5

Query Match 2.4%; Score 255; DB 3; Length 1507;  
Best Local Similarity 17.6%; Pred. No. 1.2e-10;  
Matches 164; Conservative 159; Mismatches 281; Indels 326; Gaps 35;

QY 441 VNHFOFNINVAEKLQEVFNKHFISLEOLEYSREGLVEMID-----WIDNG----- 488  
DB 673 INGNHNLNMFLEIRTVLTKYKMSNVLEPYEKSGLANHGDLSVCPSWDEBNLYKKN 732  
QY 489 -----ECIDLEKKL-----GL----- 500  
DB 733 RDYNDMCKSTWKSSTIPKMPDYKTKLLYGLYGLGRLGSLNISKVNIKFSQPNIT 792  
QY 501 LALINEES-----HPPQATDSTLLEKLSQHAN 528  
DB 793 LPMFNPSSIKNLLDLSVYCLGPLENAVNNKCFRSLFAVRHETNECIIIGTHEQERN 852  
QY 529 NH-----FYVXP-----RVAVNNFGVGVAG 549  
DB 853 NNCRTRRSDTKPNQCNVAKNISTKNWTVTSPFIRDYDEKCPRPYPLKFSFG----- 906  
QY 550 EYQYD-----VAGILEKRDTRDILN-LRESRDPYLDPEHVSNN----- 593  
DB 907 --KYDEETGCKSLINK-----NINIPLFSSCLEVNFIMYPSYLOTEKKNYGVV 958  
QY 594 -----NQDTL-----KCGSKHRRPTVSQPKVDSL-HSLMATLSSNPFVRCIKNNMK 642  
DB 959 ASESVNSSLVYAKGECCYINEKNC-----VIDKVNHRFTSLTNDIDFNQNTI--NLVK 1012  
QY 643 MBD-----QPDQAVVLNQLRYSGLMELTVRIKAGYAVARRPQDFYKRYKVMRLALPEDVR 699  
DB 1013 LDELVINNDQSSSHRAKNTPIEN-----SESTIVRKNSNAPEHR 1054  
QY 700 GKCTSLQLYDASNSWQIGKTVFRESLEQLEK--HREBVSAAVAVIRAHVIGFLA 757  
DB 1055 ---SLKINSYTPNRGENPAKESDSTRNTDESMDDEVINKREBAKNAEIRK----- 1105

QY 758 KQYKVLVYCVIIQKNYAFILR-----RPFILKKAIVFOKQAFGO-IARRVROL 811  
DB 1106 -EBAQAMAKABEERKKAELAVKABEERKRIEAKKABEERKRIEAKKABEERKRIE 1164  
QY 812 AKREOEKK---KQEEKKKQREERERERERELRAQOEETRKQOELALOKS 867  
DB 1165 AEKKAEEERKIIIAKKAEEERKRIEAKKABEERKIIIAKKAEEERKKAELAVKABE 1224  
QY 866 OKAELETRELEKOKEN-----KQVEILLEKEIEPLQRMKEQOELSL 910  
DB 1225 KKAAPAKKALTLKESRQKTIIVAVKABEERKRIEAKKABEERKRIEAKKABEERK 1284  
QY 911 TEASLOKLERRDQELRLIEERACRAQEFLESINFDIEDEVNIERSLSGSEFSSEL 970  
DB 1285 VEA-----EKKAEEERKIIIAKKAEEERIKR-----EAVKKEHEVIRKSNLSER 1331  
QY 971 AESACEKPNFNFSPQYPEEVEYDEGFADDDAFKQSPNPEHGSDDQT-----SG 1021  
DB 1332 KTAIFLKKNNNY-----ETRNIDDNSFKLDEEBYKSNINIDYRNKIISSMKEN 1381  
QY 1022 IRTSDSEBBDPYMNDTVPTSPSADSTVLLAPSVQDSGLNNSSGSSTYCPQNGDL 1081  
DB 1382 MCTNDVSSKYCDYMKDKISSGCSNDEKQLCCSISD-----YCL----- 1421  
QY 1082 PSPDDYDYPDDYEDGATSGSVTFPSNSYGSQWSPPYRCGVGTYNSSGAYRFSSEG-- 1139  
DB 1422 ---NYPDINSKYYD-----CTKKEFSD-----PLYKCSNBEYKAVY-FAGAGII 1464  
QY 1140 -----AQSFEDESEDPD 1152  
DB 1465 MSIIAICIKIIGKKMFKAFAFDEIVEDYD 1494

RESULT 10  
US-08-728-323A-2  
Sequence 2, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28, 678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-728-323A-2

Query Match 2.4%; Score 253.5; DB 2; Length 1162;  
Best Local Similarity 28.2%; Pred. No. 1e-10;  
Matches 72; Conservative 59; Mismatches 107; Indels 17; Gaps 6;

QY 795 QKQLRQIARVYRQLAKREBEKKKQEEBEKKREBEEREREREREAELRAQEE 854  
DB 753 QEQEQQEQEQEQEQEQEL-EEQEQELDQEQELEEQEQELEEQEQELEEQEQEL 811  
QY 855 TRKQELBALQSKQKAEITRELEKQENQVEILRLEKEIEDLQRMKEQELSTES 914  
DB 812 EEOEQELEEQEQELEEQE--QELEBEQELEBEQEQEQEQEQEQEQEQELEBEQ 869  
QY 915 LQKLGRRDQELRLLEEAACRAQEFLESINPDEIDECVNRNIRSLSGSSEFSSLAESA 974  
DB 870 EEOBEQEQELEBEVEEQEQELEF-VEEQEQELEBEVEEQEQEL---EEVEEQEQ 924  
QY 975 CEKPRFNSQPRPEEVEBGFADDAFKDSPNBEHSHSDORTSGITSDSSEDP- 1033  
DB 925 VEQEQETVEEPT---ILHGSSSEDEMEVDYFVVSSTHQAASSPPGDDTPDDPQGPS 980  
QY 1034 ----YNDTVVPTSP 1044  
DB 981 REYRYLRTSPPHRP 995

RESULT 11  
US-09-298-568-2  
Sequence 2, Application US/09298568  
Patent No. 632792

GENERAL INFORMATION:  
APPLICANT: Kieff, Elliott D.  
APPLICANT: Ballester, Mary E.  
APPLICANT: Kaye, Kenneth M.  
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
FILE REFERENCE: 16412-10001R  
CURRENT APPLICATION NUMBER: US/09/298,568  
CURRENT FILING DATE: 1999-04-21  
EARLIER APPLICATION NUMBER: US 60/109,422  
EARLIER FILING DATE: 1998-11-19  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1162  
TYPE: PRT  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-298-568-2

Query Match 2.4%; Score 253.5; DB 4; Length 1162;  
Best Local Similarity 28.2%; Pred. No. 1e-10;  
Matches 72; Conservative 59; Mismatches 107; Indels 17; Gaps 6;

QY 795 QKQLRQIARVYRQLAKREBEKKKQEEBEKKREBEEREREREREAELRAQEE 854  
DB 753 QEQEQQEQEQEQEQEL-EEQEQELDQEQELEEQEQELEEQEQELEEQEQEL 811  
QY 855 TRKQELBALQSKQKAEITRELEKQENQVEILRLEKEIEDLQRMKEQELSTES 914  
DB 812 EEOEQELEEQEQELEEQE--QELEBEQELEBEQEQEQEQEQEQEQEQELEBEQ 869  
QY 915 LQKLGRRDQELRLLEEAACRAQEFLESINPDEIDECVNRNIRSLSGSSEFSSLAESA 974  
DB 870 EEOBEQEQELEBEVEEQEQELEF-VEEQEQELEBEVEEQEQEL---EEVEEQEQ 924  
QY 975 CEKPRFNSQPRPEEVEBGFADDAFKDSPNBEHSHSDORTSGITSDSSEDP- 1033  
DB 925 VEQEQETVEEPT---ILHGSSSEDEMEVDYFVVSSTHQAASSPPGDDTPDDPQGPS 980

QY 1034 ----YNDTVVPTSP 1044  
DB 981 REYRYLRTSPPHRP 995

RESULT 12  
US-08-056-200-94  
Sequence 94, Application US/08056200  
Patent No. 5616500

GENERAL INFORMATION:  
APPLICANT: Steinhert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedtich, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1898 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-056-200-94

Query Match 2.3%; Score 244.5; DB 1; Length 1898;  
Best Local Similarity 26.8%; Pred. No. 1.2e-09;  
Matches 83; Conservative 64; Mismatches 96; Indels 67; Gaps 11;

QY 635 CIKPMQKMPDQFQDAVVLNQLYSGMLETVRI-RKAGYAVRPPQDFYKRYKVLNRNIA 693  
DB 196 CQHETEEPPDD-----EQLRREHLELRKRGEEKQOQRREROD-----RVFQ---- 239  
QY 694 LPEDVRGKSTSLQLYDASNSWQGLKTVFLRESLQKLEKREBEVSHAMVRAHYL 753  
DB 240 -----EEBEKEMRRRETVLRKEEKLDEEFQORELOEEBEQLRK-- 280  
QY 754 GLIARQKRVKLVYVITQKRYRFLRLRRFLHLKKAIVFOKQLRGQIARVYRQLAE 813  
DB 281 --LEQELRRRQ-----EEEQOQRLR-----EQLRRK-QEERREEQEE 320  
QY 814 KEEQEKKKQEEBEKK---KREBE-----REEREREREAELRAQEEETRKQEQELEA 863  
DB 321 RREQQERREQQERREQQRLRREQQERREQQRLRREQQERREQQRLRREQQERREQQRLR 380  
QY 864 LQSKQKAEITRELEKQENQ-VEEILRLEKEIEDLQRMKEQELSTESLQKQER 922  
DB 381 QEERREEQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 440

QY 923 DQELRLREE 932  
: : : : :  
Db 441 EQLRREOE 450

## RESULT 13

US-08-800-644-94  
; Sequence 94, Application US/08800644  
; Patent No. 5958752

## GENERAL INFORMATION:

APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/056,200

FILING DATE: 30-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Fedick, Michael P.

REGISTRATION/DOCKET NUMBER: 36,799

REFERENCE/DOCKET NUMBER: NIH054.001A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (714) 760-0404

TELEFAX: (714) 760-9502

INFORMATION FOR SEQ ID NO: 94:

SEQUENCE CHARACTERISTICS:

LENGTH: 1898 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-800-644-94

Query Match 2.3%; Score 244.5; DB 2; Length 1898;

Best Local Similarity 26.8%; Pred. No. 1.2e-09;

Matches 83; Conservative 64; Mismatches 96; Indels 67; Gaps 11;

QY 635 CLKNNQKRPDQDQVAVLNLQKRYSGMLTVRI-RKAGVAVRPDPFKRYVYLMKNTA 693

Db 196 CQGHETEPDDE-----EQLRRLRLRLRRGRREKQCORRRERD-----RVFO----- 239

QY 694 LPEDVNGKSTLSQLVYDASNSEWQLGKTYVFLRESLEQKLEKREBEVSHAAVITAHVL 753

Db 240 -----EEREKMKRKEVYLRKEEQLQEFEPKQRELQEEBQLRK----- 280

QY 754 GFLARKQYKVLVYCVIITQKNTAFLLRRFLHLKKAIVFOKLRGQIARVYRQLALAE 813

Db 281 --LEROELRRERQ-----EEOQOQRLRR-----EQLRK--QEERREQOE 320

QY 814 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

Db 321 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

QY 814 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

Db 321 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

QY 814 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

Db 321 KREOEKKKQOEBEKK--KREBE-----RERERREAVLRQOEETRKQOELTA 863

QY 864 LQSKQKAEITRELEKQENK-QVEIILRLKEIEDLQMKQOEELSTEAISLOKQERR 922  
: : : : :  
Db 381 QEERREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 440

QY 923 DQELRLREE 932  
: : : : :  
Db 441 EQLRREOE 450

## RESULT 14

US-08-910-925-1  
; Sequence 1, Application US/08910925  
; Patent No. 6162601

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,925  
FILING DATE: Herewith  
CLASSIFICATION: 514

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 514

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION/DOCKET NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0365 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 717 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: FIBRINOT01

CLONE: 53219

US-08-910-925-1

Query Match 2.0%; Score 218; DB 4; Length 717;

Best Local Similarity 23.0%; Pred. No. 2.9e-08;

Matches 125; Conservative 75; Mismatches 175; Indels 168; Gaps 23;

QY 729 LEOKLE--KREBEVSHAAVIRAHVGLARKQYRKVLY-----CVIITQKY 775

Db 209 LEOKVLAQOEWEHNAKTIK-----YIRTKYPHLYIPGRNCFATQKLIBESQRM 263

QY 776 RAFLLRRLFLHLKKAIVFOKLRGQIARVYRQLALAE-----RQOEKKKQOEBEKK 830

Db 264 NALFEGRR-----IEFADQI-NKMEARPRQSMKEKHQVYVNEBQKAEQEGKVAQ 314

QY 831 REEERERREERREAEVLRQOEETRK-----KQOELALQSKQKAEV-----TR 875

Db 315 REEELTEGNQNDVIEBAGEEERKEKALIVHSDAEKQE--EEOQKQEWVQMBEETVR 373





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:07:56 ; Search time 34.8872 Seconds

(without alignments)  
4441.378 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898  
Sequence: 1 MRLGAAVAALGRRAPAS.....INKYLEATGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3841.5	98.6	751	4 Q96T39	Q96T39 homo sapien
2	3833.5	98.3	751	4 Q96RP9	Q96RP9 homo sapien
3	3506.5	90.0	751	11 Q921D6	Q921D6 mus musculus
4	3499.5	89.8	751	11 Q921D6	Q921D6 mus musculus
5	2521.5	64.7	729	5 Q9VM33	Q9VM33 drosophila
6	2119.5	54.4	750	5 Q9XV52	Q9XV52 caenorhabditis
7	1990	51.1	757	10 Q941E1	Q941E1 oryza sativa
8	1975.5	50.7	770	10 Q941E1	Q941E1 oryza sativa
9	1972	50.6	754	10 Q96C61	Q96C61 arabidopsis
10	1970	50.5	754	10 Q9SHD6	Q9SHD6 arabidopsis
11	1964	50.4	757	10 Q9F6G4	Q9F6G4 oryza sativa
12	1947.5	50.0	757	3 Q96WU0	Q96WU0 arxula aden
13	1472	37.8	692	16 Q8YP62	Q8YP62 arabidopsis
14	1470	37.7	682	2 Q9RNG0	Q9RNG0 heliobacter
15	1427.5	36.6	700	2 Q9AF78	Q9AF78 archaebacter
16	1424.5	36.5	693	16 Q8R602	Q8R602 fusobacteri

17	1422.5	36.5	690	16 Q8RV1	Q8RV1 thermosae
18	1416.5	36.3	694	2 Q8RTE9	Q8RTE9 bartonella
19	1407	36.1	701	2 Q8RME2	Q8RME2 mycobacteri
20	1398	35.9	699	16 Q8UE15	Q8UE15 agrobacteri
21	1372.5	35.2	692	2 Q9F4B2	Q9F4B2 bacillus st
22	1367	35.1	695	16 Q9KPM5	Q9KPM5 vibrio chol
23	1365	35.0	707	2 Q9RHH6	Q9RHH6 porphyromon
24	1360	34.9	707	2 Q9RHH7	Q9RHH7 porphyromon
25	1358	34.8	707	2 Q9RHH6	Q9RHH6 porphyromon
26	1356	34.8	707	2 Q9RHH4	Q9RHH4 porphyromon
27	1354	34.7	707	2 Q9RHH5	Q9RHH5 porphyromon
28	1351.5	34.7	558	3 Q9U0A3	Q9U0A3 schizobacch
29	1341.5	34.4	783	10 Q9S175	Q9S175 arabidopsis
30	1337.5	34.3	689	2 Q9VMU0	Q9VMU0 mycoplasma
31	1334.5	34.2	783	10 Q94BR7	Q94BR7 arabidopsis
32	1142.5	29.3	681	2 Q9A1G7	Q9A1G7 carsonella
33	1074.5	27.6	696	16 Q8XPE6	Q8XPE6 clostridium
34	1074	27.6	700	16 Q8R7R5	Q8R7R5 thermosae
35	1015.5	26.1	779	4 Q969S9	Q969S9 homo sapien
36	988	25.3	777	11 Q8R2Q4	Q8R2Q4 mus musculu
37	959	24.6	732	4 Q8WY10	Q8WY10 homo sapien
38	921.5	23.6	690	16 Q8R604	Q8R604 fusobacteri
39	803.5	20.6	678	16 Q9RXC2	Q9RXC2 deinoxoccus
40	786.5	20.2	813	3 Q944Z9	Q944Z9 schizobacch
41	784.5	20.1	1093	5 Q9VCX4	Q9VCX4 drosophila
42	754	19.3	689	5 Q95Y73	Q95Y73 caenorhabdi
43	737	18.9	694	16 Q8YPI9	Q8YPI9 arabidopsis
44	709	18.2	513	4 Q9H6Z1	Q9H6Z1 homo sapien
45	700	18.0	236	5 Q00887	Q00887 leishmania

## ALIGNMENTS

RESULT 1	ID	Q96T39	PRELIMINARY	PRT	751 AA
AC	Q96T39	01-DEC-2001 (TRENBLREL. 19, Created)			
DT	01-DEC-2001 (TRENBLREL. 19, Last sequence update)				
DT	01-JUN-2002 (TRENBLREL. 21, Last annotation update)				
DE	Elongation factor G1.				
OS	EFGL.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OK	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Hammarund M., Wilson W., Corcoran W., Binhorn S., Grander D.,				
RA	Sangfelt O.,				
RT	"Identification and characterization of two novel human elongation				
RT	factor genes, hefg2 and hefg1, phylogenetically conserved through				
RT	evolution."				
RL	Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AF367998; AAK53402.1; -				
DR	InterPro; IPR004540; EF-G.				
DR	InterPro; IPR000640; EF-G.				
DR	InterPro; IPR004161; EFTU_D2.				
DR	InterPro; IPR000795; EF_GTPbind.				
DR	InterPro; IPR005225; Small_GTP.				
DR	Pfam; PF00673; EFG_G_1.				
DR	Pfam; PF00009; GTP_EFTU_1.				
DR	Pfam; PF03144; GTP_EFTU_D2_1.				
DR	TIGRFAMs; TIGR00484; EF-G_1.				
DR	TIGRFAMs; TIGR00231; Small_GTP_1.				
DR	PROSITE; PS00301; EFATOR GTP; UNKNOWN_1.				
KW	GTP-binding; Protein biosynthesis.				
SO	SEQUENCE 751 AA; 83471 MW; 5937FFB2A089E2E CRC64;				
Query Match	98.6%;	Score 3841.5;	DB 4;	Length 751;	
Best Local Similarity	98.8%;	Pred. No. 2.3e-258;			
Matches 743;	Conservative	2;	Mismatches 6;	Indels 1;	Gaps 1;

```

QY 1 MRLGAAVAALGRGRAPASLGMQRKOVNMKACRMSSSGVIPNEKIRNIGISAHIDSXT 60
DB 1 MRLGAAVAALGRGRAPASLGMQRKOVNMKACRMSSSGVIPNEKIRNIGISAHIDSXT 60
QY 61 TLTERVLYYTGRIAKMHEVKGKDGAVAVMDSMELEROGITTIQSAATFTMMKDVNINIID 120
DB 61 TLTERVLYYTGRIAKMHEVKGKDGAVAVMDSMELEROGITTIQSAATFTMMKDVNINIID 120
QY 121 TPGHVDFTIEVERALRVLDAVAVLCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
DB 121 TPGHVDFTIEVERALRVLDAVAVLCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
QY 181 SNPARALQOMRSLKNHNTAFMOJIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
DB 181 SNPARALQOMRSLKNHNTAFMOJIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
QY 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
QY 301 LKNKGVOPLDVALEYLPPNSEVQNYAAILNKDSEKTKIIMNSSRNSHPVGLAPFL 360
DB 301 LKNKGVOPLDVALEYLPPNSEVQNYAAILNKDSEKTKIIMNSSRNSHPVGLAPFL 360
QY 361 EVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLARHADMMEASTEEVYAGDICA 420
DB 361 EVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLARHADMMEASTEEVYAGDICA 420
QY 421 LFGIDCAGSDTFTDKANSGLSMESIHPDPVSIAMKPSNKNDLEKFSKIGRFTREDPT 480
DB 421 LFGIDCAGSDTFTDKANSGLSMESIHPDPVSIAMKPSNKNDLEKFSKIGRFTREDPT 480
QY 481 FKYPFDTEKETEYISGMGELHLEIYAQRLEBYGCPCTIGKPKVARETITAPVPDFTH 540
DB 481 FKYPFDTEKETEYISGMGELHLEIYAQRLEBYGCPCTIGKPKVARETITAPVPDFTH 540
QY 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSNIPKQFPAVEKGFADACKGFL 600
DB 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSNIPKQFPAVEKGFADACKGFL 600
QY 601 SGHKSGLRFLVODGAHHMVDNSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
DB 601 SGHKSGLRFLVODGAHHMVDNSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
QY 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752
DB 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752

RESULT 2
Q96RP9 PRELIMINARY; PRT; 751 AA.
AC 096RP9;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DB 01-DUN-2002 (TEMBLrel. 21, Last annotation update)
DB Elongation factor G.
GN EFG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21369185; PubMed=11374907;
RT Cloning and Characterization of Human and Mouse Mitochondrial
RT Elongation Factor G, Gfm and Gfm, and Mapping of Gfm to Human

```

```

RT Chromosome 3q25.1-q26.2.;
RL Genomic 74:109-114(2001).
DR EMBL; AF309777; AAK56877.1; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG-C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C_1.
DR Pfam; PF00009; GTP_EFTU_1.
DR Pfam; PF03144; GTP_EFTU_D2_1.
DR TIGRPFAM; TIGR00484; EF-G_1.
DR TIGRPFAM; TIGR00231; small_GTP_1.
DR PROSITE; PS00301; EFATOR_GTP; UNKNOWN_1.
KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 751 AA; 83505 MW; BF794C925A6775C2 CRC64;

```

```

Query Match 98.3%; Score 3833.5; DB 4; Length 751;
Best Local Similarity 98.5%; Pred. No. 8.2e-258;
Matches 741; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

```

```

QY 1 MRLGAAVAALGRGRAPASLGMQRKOVNMKACRMSSSGVIPNEKIRNIGISAHIDSXT 60
DB 1 MRLGAAVAALGRGRAPASLGMQRKOVNMKACRMSSSGVIPNEKIRNIGISAHIDSXT 60
QY 61 TLTERVLYYTGRIAKMHEVKGKDGAVAVMDSMELEROGITTIQSAATFTMMKDVNINIID 120
DB 61 TLTERVLYYTGRIAKMHEVKGKDGAVAVMDSMELEROGITTIQSAATFTMMKDVNINIID 120
QY 121 TPGHVDFTIEVERALRVLDAVAVLCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
DB 121 TPGHVDFTIEVERALRVLDAVAVLCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
QY 181 SNPARALQOMRSLKNHNTAFMOJIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
DB 181 SNPARALQOMRSLKNHNTAFMOJIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
QY 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDLKLAIRATLKRSFTPVFLGSA 300
QY 301 LKNKGVOPLDVALEYLPPNSEVQNYAAILNKDSEKTKIIMNSSRNSHPVGLAPFL 360
DB 301 LKNKGVOPLDVALEYLPPNSEVQNYAAILNKDSEKTKIIMNSSRNSHPVGLAPFL 360
QY 361 EVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLARHADMMEASTEEVYAGDICA 420
DB 361 EVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLARHADMMEASTEEVYAGDICA 420
QY 421 LFGIDCAGSDTFTDKANSGLSMESIHPDPVSIAMKPSNKNDLEKFSKIGRFTREDPT 480
DB 421 LFGIDCAGSDTFTDKANSGLSMESIHPDPVSIAMKPSNKNDLEKFSKIGRFTREDPT 480
QY 481 FKYPFDTEKETEYISGMGELHLEIYAQRLEBYGCPCTIGKPKVARETITAPVPDFTH 540
DB 481 FKYPFDTEKETEYISGMGELHLEIYAQRLEBYGCPCTIGKPKVARETITAPVPDFTH 540
QY 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSNIPKQFPAVEKGFADACKGFL 600
DB 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSNIPKQFPAVEKGFADACKGFL 600
QY 601 SGHKSGLRFLVODGAHHMVDNSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
DB 601 SGHKSGLRFLVODGAHHMVDNSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
QY 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752
DB 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752

```

RESULT 3  
ID 092106 PRELIMINARY; PRT: 751 AA.  
AC 092106;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Unknown (protein for MG:7961).  
GN GFM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Straube R. J.  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC013093; AAH1093.1; -.  
DR MGD; MG1:107339; Gfm.  
DR InterPro; IPR004540; EF-G.  
DR InterPro; IPR000640; EF-G.  
DR InterPro; IPR004161; EFTU D2.  
DR InterPro; IPR000795; EF\_GTPbind.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00679; EFG\_C\_1.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR Pfam; PF03144; GTP\_EFTU D2; 1.  
DR TIGRFAMs; TIGR00484; EF-G; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR PROSITE; PS00301; EFACITOR\_GTP; UNKNOWN\_1.  
KW GTP-binding; Protein biosynthesis.  
SQ SEQUENCE 751 AA; 83539 MW; 1DCE4D8981EAD2C CRC64;

Query Match 90.0%; Score 3506.5; DB 11; Length 751;  
Beet Local Similarity 89.4%; Pred. No. 4.4e-235;  
Matches 675; Conservative 34; Mismatches 39; Indels 7; Gaps 4;

1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 57  
1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 57  
1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 57  
58 GKTLLTERVLYYTGRIAMHEVKGKDGAVVNDSELERQGITTSAAATFTMKDWIN 117  
59 GKTLLTERVLYYTGRIATMHEVKGKDGAVVNDSELERQGITTSAAATFTMKDWIN 118  
118 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQMKRYNVPFLTFINKD 177  
119 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQMKRYNVPFLTFINKD 178  
178 RMGSPARALQOMRSKLNHNTAFMQIPMGLGNPFGIVDLIERAIYFDGDSQIVRYGE 237  
179 RMGSPARALQOMRSKLNHNTAFMQIPMGLGNPFGIVDLIERAIYFDGDSQIVRYGE 238  
238 IPAEALRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDKLAIIRATLKRSFTPYFL 297  
239 IPAGRAAADAHRQELIECVANSDEQLGEMPLEEKIPISIDKLAIIRATLKRSFTPYFL 298  
238 GSALNKQGVQPLLDVAVLEVLPNPEVQNYAIIANKKDSSEKTKILMNSRRHNSHPVGIA 357  
239 GSALNKQGVQPLLDVAVLEVLPNPEVQNYAIIANKKDSSEKTKILMNSRRHNSHPVGIA 357  
358 PFLVGRFGQLTVYVSYOGELKKGDTIYTRTRKKYVRLRLARMADMMEAETEEVYAD 417  
359 PFLVGRFGQLTVYVSYOGELKKGDTIYTRTRKKYVRLRLARMADMMEAETEEVYAD 417  
358 PFLVGRFGQLTVYVSYOGELKKGDTIYTRTRKKYVRLRLARMADMMEAETEEVYAD 416  
418 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 477  
419 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 477  
417 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 476  
478 DPTFKVYPTENKEVYISGMGELHLEIYAKQLEBEVGCITGKPVNARETITAPVPD 537  
479 DPTFKVYPTENKEVYISGMGELHLEIYAKQLEBEVGCITGKPVNARETITAPVPD 537  
477 DPTFKVYPTENKEVYISGMGELHLEIYAKQLEBEVGCITGKPVNARETITAPVPD 536

538 FTHKKQSGAGQYGVKIVLEPLDEPDYTKLEFSDTEFGSNI PKQFPAVEKGFIDACEK 597  
539 FTHKKQSGAGQYGVKIVLEPLDEPDYTKLEFSDTEFGSNI PKQFPAVEKGFIDACEK 596  
537 FTHKKQSGAGQYGVKIVLEPLDEPDYTKLEFSDTEFGSNI PKQFPAVEKGFIDACEK 596  
598 GPLSGHKLSGLRFVYQDGAHNMVDSNEISFIRAGEGALKOLANATCILPEINAVEVYA 657  
599 GPLSGHKLSGLRFVYQDGAHNMVDSNEISFIRAGEGALKOLANATCILPEINAVEVYA 656  
658 PNEFGQVYIAGINRHSVITGDDVEDYFTLYADVPVLDMMGYSFELASCTEGKGEYTM 717  
659 PNEFGQVYIAGINRHSVITGDDVEDYFTLYADVPVLDMMGYSFELASCTEGKGEYTM 716  
657 PNEFGQVYIAGINRHSVITGDDVEDYFTLYADVPVLDMMGYSFELASCTEGKGEYTM 716  
718 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 752  
719 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 751  
717 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 751

RESULT 4  
ID 092410 PRELIMINARY; PRT: 751 AA.  
AC 092410;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Elongation factor G.  
GN GFM OR EFG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=21269185; PubMed=11374907;  
Gao J., Yu L., Zhang P., Jiang J., Chen J., Peng J., Wei Y., Zhao S.;  
"Cloning and Characterization of Human and Mouse Mitochondrial  
RT Elongation Factor G, GFM and Gfm, and Mapping of Gfm to Human  
RT Chromosome 3q25.1-q26.2.";  
RL Genomics 74:109-114(2001).  
DR EMBL; AF15511; AAKS8878.1; -.  
DR MGD; MG1:107339; Gfm.  
DR InterPro; IPR004540; EF-G.  
DR InterPro; IPR000640; EF-G.  
DR InterPro; IPR004161; EFTU D2.  
DR InterPro; IPR000795; EF\_GTPbind.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00679; EFG\_C\_1.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR Pfam; PF03144; GTP\_EFTU D2; 1.  
DR TIGRFAMs; TIGR00484; EF-G; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR PROSITE; PS00301; EFACITOR\_GTP; UNKNOWN\_1.  
KW GTP-binding; Protein biosynthesis.  
SQ SEQUENCE 751 AA; 83667 MW; C206377C48E97808 CRC64;

Query Match 89.8%; Score 3499.5; DB 11; Length 751;  
Beet Local Similarity 89.1%; Pred. No. 1.4e-234;  
Matches 673; Conservative 36; Mismatches 39; Indels 7; Gaps 4;

1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 57  
1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 57  
1 MRLGAAVAALGRG---RAPASLGMRQVNMKACRMSSSGVIPNEKIRNIGISAHIDS 58  
58 GKTLLTERVLYYTGRIAMHEVKGKDGAVVNDSELERQGITTSAAATFTMKDWIN 117  
59 GKTLLTERVLYYTGRIATMHEVKGKDGAVVNDSELERQGITTSAAATFTMKDWIN 118  
118 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQMKRYNVPFLTFINKD 177  
119 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQMKRYNVPFLTFINKD 178  
178 RMGSPARALQOMRSKLNHNTAFMQIPMGLGNPFGIVDLIERAIYFDGDSQIVRYGE 237

Db 179 RMGNSRRLQOMRSKLNHNAAFVQIPIGEGDFKGIIDLIBERRAIFPDGFCQIVRYDB 238  
 Qy 238 IPAEIAPAAATDHRCEIIECVANSDEQGEFLEBKIPISIDLTALIRATLKSPFVPL 297  
 Db 239 IPAGLAAAADHROELIECVANSDEQGEFLEBKIPISIDLTALIRATLKSPFVPL 298  
 Qy 298 GSALKNKGVQPLLDVAVLEYPNPESEVONYAILNKDSEKERTILMNSRHNSHPVGLA 357  
 Db 299 GSALKNKGVQPLLDVAVLEYPNPESEVONYAILNKDSEKERTILMNSRHNSHPVGLA 357  
 Qy 358 FPLEVRFQGLTYVRSYQGLKGGDTTYNTRTKKRVRLQRLAMADHMASTIEEYVAD 417  
 Db 358 FPLEVRFQGLTYVRSYQGLKGGDTTYNTRTKKRVRLQRLAMADHMASTIEEYVAD 416  
 Qy 418 ICALFGIDCAGDTFIDKANGSGMESIHVPDPVSIAMKPSKNDLEKSKIGRTBE 477  
 Db 417 ICALFGIDCAGDTFIDKANGSGMESIHVPDPVSIAMKPSKNDLEKSKIGRTBE 476  
 Qy 478 DPTFKVYPTENKEVYISGMGLHLEIYAQRLEREXGCPCTITGKPRVARETTAPVPD 537  
 Db 477 DPTFKVHFPDESKETIVSGMGLHLEIYAQRMEREXGCPCTITGKPRVARETTAPVPD 536  
 Qy 538 FTHKSGGAGQGVKIVLEPLDPEDTYLERSDFTGSGNIPKQVPAVEKFLDACCK 597  
 Db 537 FTHKSGGAGQGVKIVLEPLDPEDTYLERSDFTGSGNIPKQVPAVEKFLDACCK 596  
 Qy 598 GPLSGHKLSCGLRVLDGAMHWDSNEISITRAGEGALKOALNATLCTILEPIMAVEVYA 657  
 Db 597 GPLSGHKLSCGLRVLDGAMHWDSNEISITRAGEGALKOALNATLCTILEPIMAVEVYA 656  
 Qy 658 PNEFGQVIVAGINRRHGVITGQGVEDYFTLVADVPDPMFGYSTELRSCTECKGEYME 717  
 Db 657 PNEFGQVIVAGINRRHGVITGQGVEDYFTLVADVPDPMFGYSTELRSCTECKGEYME 716  
 Qy 718 YSRVOPCLSTQEDVINKYLEATGOLPYKKGKAKN 752  
 Db 717 YSRVOPCLSTQEDVINKYLEATGOLPYKKGKAKN 751  
 RESULT 5  
 Q9VM33 PRELIMINARY; PRT; 729 AA.  
 AC Q9VM33; 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE CG4567 protein.  
 GN CG4567.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adam M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Beeson A., Bakendole J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokhov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.B., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harts N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Krafc C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reihert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svetkae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195 (2000).  
 DR EMBL: AE003616; AAF52495.1; -.  
 DR HSSP: P13511; IFFM.  
 DR Flybase: FBgn0031898; CG4567.  
 DR InterPro: IPR004540; EF-G.  
 DR InterPro: IPR000640; EFG-C.  
 DR InterPro: IPR004161; EFTU D2.  
 DR InterPro: IPR000785; EF-GTbind.  
 DR InterPro: IPR005225; Small-GTP.  
 DR Pfam: PF00679; EFG\_C\_1.  
 DR Pfam: PF00009; GTP-BPTU\_1.  
 DR Pfam: PF03144; GTP-BPTU D2; 1.  
 DR PRINTS: PR00315; E1ONGATNFC.  
 DR TIGRFAMS: TIGR00484; EF-G\_1.  
 DR TIGRFAMS: TIGR00231; small-GTP\_1.  
 DR TIGRFAMS: TIGR00301; EFACOR-GTP\_1.  
 DR PROSITE: PS00301; EFACOR-GTP\_1.  
 DR GTP-binding: Protein biosynthesis.  
 SQ SEQUENCE 729 AA; 81739 MW; 6CB15A76DF1807AB CRC64;  
 Query Match 64.7%; Score 2521.5; DB 5; Length 729;  
 Best Local Similarity 66.1%; Pred. No. 1.2e-166;  
 Matches 469; Conservative 101; Mismatches 136; Indels 3; Gaps 2;  
 Qy 44 EKIRNIGISAHIDSQKTLTERLVLYYGRISIAHBEVKGAGVAMDSMELEKRGITIQ 103  
 Db 24 KIRNIGISAHIDSQKTLTERLVLYYGRISIAHBEVKGAGVAMDSMELEKRGITIQ 83  
 Qy 104 SAATFTMKDVNINIIDTPGHVDTIEVERALRVLDGAVLVCAVGVQCQTVMYRNQK 163  
 Db 84 SAATFTMKDVNINIIDTPGHVDTIEVERALRVLDGAVLVCAVGVQCQTVMYRNQK 143  
 Qy 164 RYNVPEFLFINKLDMGNSPAPALQOMRSKLNHNAFQIPMGLEGNFRGIYDLLEBAI 223  
 Db 144 RYNVPEFLFINKLDMGNSPAPALQOMRSKLNHNAFQIPMGLEGNFRGIYDLLEBAI 203  
 Qy 224 YDQGSQIVRGEIPAEIAPAAATDHRCEIIECVANSDEQGEFLEBKIPISIDLTAL 283  
 Db 204 YEGEHGMDIRDELIPDMKRVSELERQELIEHLSNABETLGELTLEKGFREDIDIKAL 263  
 Qy 284 RRATLKRSFTPVFGSALKNKGVQPLLDVAVLEYPNPESEVONYAILNKDSEKERTILM 343  
 Db 264 RRATLKRSFTPVFGSALKNKGVQPLLDVAVLEYPNPESEVONYAILNKDSEKERTILM 321  
 Qy 344 NSSRNSHPFVGLAPPLEVGRGQGLTYVRSYQGLKGGDTTYNTRTKKRVRLQRLAM 403  
 Db 322 NPARGKDPFVGLAPPLEVGRGQGLTYVRSYQGLKGGDTTYNTRTKKRVRLQRLAM 381  
 Qy 404 DWKEASTIEVYVGDICAFGIDCAGDTFIDKANGSGMESIHVPDPVSIAMKPSKND 463  
 Db 382 NQME-DVNEVYAGDIFALFGVDCAGDTFTTPPKNNLSMESIFVEPVPVSAIKRNTD 440  
 Qy 464 LEKFGKIGRFTREDPTFKVYPTENKEVYISGMGLHLEIYAQRLEREXGCPCTITGK 523

Db	441	RDNFBKALARKFKEDPTTFPFEDNDVKETLVSMSGLHLEIYQNRNREYGCYVTLGKFK	500
Qy	554	VAFRETTIAPVPDFDTHKKQSGAGQYKVIQVLEPLDPEDEVTKLEPSDFTGSGNTPKOF	583
Db	501	VAFRETLVGPCEFDYLAHKQSGSGGQYARIIGMEPLPPNQNTLFFVETVETVQNVKOP	560
Qy	584	VPAVKGFLDACEKEPGLSGKLSGLRFLVQODGAAHVDSNIEISFIRAGEGALKQALANAT	643
Db	561	VPGVKGYRHEMAEKGMLSGHKLSTGIRFLQDGGHHIVDSNELAFMLAAGAIKEVPQNS	620
Qy	644	LCILIEPIAAVEVAPRPEQGVYAGINRNRHGVITGDDGVDEVTLTADVPILNPGSYSE	703
Db	621	WOIEPIHMLVETAPBEFOGAVMGHSKSHGIIITGEGTBGMVTVAEVPFLNMFAGH	680
Qy	704	LRSCTEGKGEYMEYSRYOPLPSTOEDVINKYLEATGOJLVKKGAKN	752
Db	681	LRSSTOGKEBFTMEYSRSPCLPDVDDQTLVROVQEOGCLAQDPKKKKN	729

RESULT 6		
ID	09XV52	
AC	09XV52	PRELIMINARY;      PRT;      750 AA.
DT	01-NOV-1999	(TrEMBLrel. 12, Created)
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)
DE	P29C12.4 protein.	
GN	P29C12.4.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae	
OC	Rhabditidae; Pelodierinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Steward C.A.;	
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=99069613; PubMed=9851916;	

RT "Genome sequence of the nematode *C. elegans*: A platform for  
investigating biology".  
RL Science 282:2012-2018 (1998).  
DR EMBL, Z81519; CAB04216.1; -.  
DR HSSP; P13551; 1FNM.  
DR InterPro; IPR004540; EF-G.  
DR InterPro; IPR000640; EF-G\_C.  
DR InterPro; IPR004161; EPTU D2.  
DR InterPro; IPR00785; EF-GTPbind.  
DR InterPro; IPR005225; Small\_GTP.  
DR Pfam; PF00679; EF-G\_C; 1.  
DR Pfam; PF00009; GTP\_EPTU; 1.  
DR Pfam; PF03144; GTP\_EPTU D2; 1.  
DR PRINTS; PR00315; BLONGATNCT.  
DR TIGRFAMs; TIGR00484; EF-G; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR PROSITE; PS00301; EFACTOR\_GTP; 1.  
KW GTP-binding; Protein biosynthesis.  
KW SEQUENCE 750 AA; 83654 MW; D054528698A18ABD CRC64;

	Query Match	Similarity	Score	DB 5	Length	750
Best Local	414	Conservative	54.9%	Pred No. 1	1e-138	
Matches	414	Conservative	125	Mismatches	198	Indels 17, Gaps 6
Qy	9	VALLGRAPASLQWQKQVNNKACWSS	-----	GYPNKIRNIGISAHIDSG	59	
Db	1	MSPLRPSVAAC--SRQLNVLRRFPAENEA	SSVYVPGVRIETIRNIGISAHIDSG	57		
Qy	60	TTLTERVLYYTGRIAMHEVKGKDGAV	MSDELRCGIIITQSATFTMMKDVINII	119		
Db	58	TTVTRILIVAPRIISMHEVRGKDVGM	TFMDLRCGIIITQSATVYDMHGNTINII	117		

```

QY 120 DTGCHVDFTIEVERALRVLTDGAVVYLCAVGVCQOCTMYRQMKRYVNPPLTPIKLDQM 179
Db 118 DTRGHVDFTIEVERALRVLTDGAVVYLCAVGVCQOCTPYVNOQLARVNPPLCFPNKQDRN 177
QY 180 GSNPARALOQMSKJLNHTAFMOQIPMGLEGNPKGVLDIEERATYFQODFSQIARYGEIP 239
Db 178 GAFPLKALDGLRKLNHNAALLHLPIGKDSNPNNGVVDLVEGHALYIEGEGLLYRDOEIP 237
QY 240 AELPAAATDHRQELIECVANSDEQLGEMFLEBKIPISIDLKLAIRLTLKRSPTPVFGS 299
Db 238 KDLRVAEADRRQELIEHIANVDELTGEMFELNDQTPNVOQIHEAIRRYVKAFPVYVLSGS 297
QY 300 ALKNKQVOPLDVAVLEYLTPNSEYONVYALINKKDSKEKTKILNNSBRHNSHPVGLAPF 359
Db 298 ALKNKQGTQMTNSVVKLIPDPSEVNRAITVKTETTGDBK-GILSPKQNDKXPFVGLAPF 356
QY 360 LEVGRFGOULTYVRSYOGELBKGDITYNTRTRKRYALQLAHMH-ADIMEASTEEVYAGDI 418
Db 357 LEAGKTKGOLYPRVYVGOQLSKGDIYASDRGSKRVQGLVYMHAADEBITT--AVAGDI 414
QY 419 CALPGIDCASGDTFTDKANSGLSMSIHVPDPVVISIANKPSNKDLEKFSIGIORFRED 478
Db 415 CAFFGJGDCHSGEFTFSDONLAPHCESMAIPEBVISMALKPVRADADNFIATLRFPKED 474
QY 479 PTFKYVYDENKKTIVSGMGEHLLEIYVORLEREYGCPCITGKPKVAFRETTIYVPPDF 538
Db 475 PTFRREBNQAKETIYVSGMGEHLLEIYVQRMKSEBNCVELGKPTVAYRBCLSGPFYFHF 534
QY 539 THKKQSGAGQYKAVIGVLEPLDPBEDYTKLEBSEDTFGSNSI PKQVPVAVEKGFPLDACEG 598
Db 535 RHKKQGTGQOQOFEIETGVIDPLPSRNTVVEFSDETFGNNIPKULFALPKKGDAIYVAG 594
QY 599 PLSGHKLSGARFVLDQGAHMHWDNSEIPIRAGBALQALANNTICILBPIYMAVEVPAR 658
Db 595 PLIKSRIRAGIHARIQDGSHTAVDSSTEIMINTQMMKSESEKANWILLBETIMVKEZTPV 654
QY 659 NEFGQVATIGINRRHGVITGQDGVDEFTLYADVPYLMFGYSTELRSCTEGKGEYTMEX 718
Db 655 TBEQGVNVVSLTGQRNALITTTDSTEGATVVICAPLMSFPGYTSBIALSLTEGKGFEFMEY 714
QY 719 SRVQPCPLSTQEDVYINKYLEATGOL-PVKGKAK 751
Db 715 SRVAPPTLEADQVQAEWQOLHGIDAPNEKGGKK 748

```

```

RESULT 7
094LEI
ID 094LEI PRELIMINARY; PRT; 757 AA.
AC 094LEI;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mitochondrial elongation factor G.
DE
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzeae; Oryza.
OX NCBI_TaxId=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Bueli C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,
RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tselirin T.,
RA Riggs F., Hsiao J., Zisemann V., Blunt S., Pai G., Vankken S.B.,
RA Unterbach T.R., Feldblum T.V., Quackenbush J., Salzberg S.L.,
RA White O., Frazer C.M.;
RT "Oryza sativa chromosome 3 BAC OSUNBa0026A15 genomic sequence.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084404; AAKS0578.1; -
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.

```

DR InterPro: IPR005225; Small\_GTP;  
 DR Pfam: PF00679; EFG\_C; 1.  
 DR Pfam: PF00009; GTP\_EFTU; 1.  
 DR Pfam: PF03144; GTP\_EFTU\_D2; 1.  
 DR TIGRfam: TIGR00464; EF-G; 1.  
 DR TIGRfam: TIGR00231; small\_GTP; 1.  
 DR PROSITE: PS00301; EFATOR\_GTP; UNKNOWN 1.  
 DR Elongation factor; GTP-binding; Protein biosynthesis.  
 KW Elongation factor; GTP-binding; Protein biosynthesis.  
 SQ SEQUENCE 757 AA; 83924 MW; C3E7C7068ACD7B93 CRC64;

Query Match 51.1%; Score 1990; DB 10; Length 757;  
 Best Local Similarity 53.6%; Pred. No. 1,2e-129;  
 Matches 396; Conservative 120; Mismatches 209; Indels 14; Gaps 6;

3 LIGAAVAALGRAPASLGMQRKQVNMKACRWSSSGVYIPEKINIGISAHIDSGKTL 62  
 LSAAAAAAAAARGMSSASA--LRARDEKEVARRES---MDMRNIGISAHIDSGKTL 84  
 63 TERVLYTGRIAKMEVKGKGVGVAVMDSMELEORGITIOSAATFTMMKDVNIIDTP 122  
 TERVLYTGRHIEHVEGRDGVGAKMDSMDLEREKGITIOSAATFTMMKDVNIIDTP 144  
 123 GHVDFTEVERALRVLDGAVLVLCVGVQOCMTVNRQMKRYNVPFTFINKLDRMGSN 182  
 GHVDFTEVERALRVLDGAVLVLCVGVQOSITVDROMRYEIPRAVAFINKLDRMGAD 204  
 145 GHVDFTEVERALRVLDGAVLVLCVGVQOSITVDROMRYEIPRAVAFINKLDRMGAD 204  
 183 PARALQOMSKLNHTAFMQIPMGLEGNFGIVDLIEERATVPDGFSGQIVRGESIPAE 242  
 PKVLNQVLSGAKLRHNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVVADVPNM 264  
 205 PKVLNQVLSGAKLRHNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVVADVPNM 264  
 243 RAAATDHROELIECVANSDEQLGEMPLEEKIPISIDKLAIIRATLKRSFTFVLSALK 302  
 265 ODVMEKREELIEVSEVDQLAEAFLENDPIQANQLKAIRATVARKETIPVMSARFK 324  
 303 NKGVQLDVLAVLEVLPSEVQVYALINKDDSEKTKILMSSRNNSHPFVGLAPLEY 362  
 325 NKGVQLDVLAVLEVLPSEVQVYALINKDDSEKTKILMSSRNNSHPFVGLAPLEY 378  
 363 GREGLTYVRSYOGELKKGDTIYNTRTKKVRQLRALRHADMEASTEVEVAGDICALF 422  
 379 GREGLTYLRIVGYVRKGDPIYNNVTGKKIKPRLVRHNSME-DIQENAHGQIVAVF 437  
 423 GIDCASGDTFTDKANSGLSMESIHVPDVISIAMKPSNNKDLKFSKIGRTREDPTK 482  
 438 GVDCAAGDTFTD-GSVYKWTSMNVBEPMVSLAVSDISKSGQFSAKALNRFQKEDPTR 496  
 483 VYEDTNKKEVTVSGMEHLHETIAORLEBYGECCTGPKVAFRETTIAPVFDTHKK 542  
 497 VGIDPESGETTISGMELHLDIYVERIRREYKVDKAVKPRVNFRETTIQRAEFDLHK 556  
 543 QSGGAGQYGVKIVGLPELPDEDTKLEFSDFTFGSNIPKQFVAVKEGFLDACCEKPLSG 602  
 557 QSGGAGQYGVKIVGLPELPDEDTKLEFSDFTFGSNIPKQFVAVKEGFLDACCEKPLSG 616  
 603 HKLSGLRFLVQDGAHNMVDSNEISFRAGEGALKQALANATLCILEPMAVEVADNEFO 662  
 617 HVENIRIRIVTQDASHAVDSSELAFLKASYAFRCQAARPVILEPVMKVELKPTERO 676  
 663 GOVIATINRHGVITQDGVEDFTLYADVPLNDMGYSTELASCEBEGEYMEISRQ 722  
 677 GTVTDGMRKKGIIIVNDGEGDTPVAVCHVPLNMGYSTALSKMTQKGEEFMEYLEIN 736  
 723 PCLPSTQEDVINKYLEATG 741  
 737 TVSGDVQMQVLVNTYKASRG 755

RESULT 8

094151 PRELIMINARY; PRT; 770 AA.

AC 094151; 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 OS Mitochondrial elongation factor G.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; Oryza.  
 OX NCBI\_TaxID=4530;

RP SEQUENCE FROM N.A.  
 RA Huang E.N., de la Bastide M., Preston R.R., Nascimben L.U.,  
 RA Spiegel L.A., Vil M.D., Baker J.P., Shah R.S., Kirchhoff K.A.,  
 RA Rodriguez M.A., King L., O'Shaughnessy A., Bal H., Dedhia N.N.,  
 RA McCombie W.R.;  
 RT "Genomic Sequence For Oryza sativa, Nipponbare strain, Clone  
 RT OSUB0084C09, Complete Sequence";  
 RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA McCombie W.R.;  
 RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Palmer L.E., Bal H., Huang E.N., de la Bastide M., Preston R.R.,  
 RA Nascimben L.U., Spiegel L.A., Vil M.D., Baker J.P., Shah R.S.,  
 RA Kirchhoff K.A., Rodriguez M.A., King L., O'Shaughnessy A., Dedhia N.N.,  
 RA McCombie W.R.;  
 RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Palmer L.E., Bal H., Huang E.N., de la Bastide M., Preston R.R.,  
 RA Nascimben L.U., Spiegel L.A., Vil M.D., Baker J.P., Shah R.S.,  
 RA Kirchhoff K.A., Rodriguez M.A., King L., O'Shaughnessy A., Dedhia N.N.,  
 RA McCombie W.R.;  
 RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC016781; AK53868.1; --

DR InterPro: IPR004540; EF-G.

DR InterPro: IPR00640; EFG\_C.

DR InterPro: IPR004161; EFTU\_D2.

DR InterPro: IPR000795; EF\_GTPbind.

DR InterPro: IPR005225; Small\_GTP.

DR Pfam: PF00679; EFG\_C; 1.

DR Pfam: PF00009; GTP\_EFTU; 1.

DR Pfam: PF03144; GTP\_EFTU\_D2; 1.

DR TIGRfam: TIGR00464; EF-G; 1.

DR TIGRfam: TIGR00231; small\_GTP; 1.

DR PROSITE: PS00301; EFATOR\_GTP; UNKNOWN 1.

DR Elongation factor; GTP-binding; Protein biosynthesis.

KW Elongation factor; GTP-binding; Protein biosynthesis.

SQ SEQUENCE 770 AA; 85268 MW; 94DD95F5119606A CRC64;

Query Match 50.7%; Score 1975.5; DB 10; Length 770;  
 Best Local Similarity 52.7%; Pred. No. 1,2e-128;  
 Matches 396; Conservative 121; Mismatches 208; Indels 27; Gaps 7;

3 LIGAAVAALGRAPASLGMQRKQVNMKACRWSSSGVYIPEKINIGISAHIDSGKTL 62  
 LSAAAAAAAAARGMSSASA--LRARDEKEVARRES---MDMRNIGISAHIDSGKTL 84  
 63 TERVLYTGRIAKMEVKGKGVGVAVMDSMELEORGITIOSAATFTMMKDVNIIDTP 122  
 TERVLYTGRHIEHVEGRDGVGAKMDSMDLEREKGITIOSAATFTMMKDVNIIDTP 144  
 123 GHVDFTEVERALRVLDGAVLVLCVGVQOCMTVNRQMKRYNVPFTFINKLDRMGSN 182  
 GHVDFTEVERALRVLDGAVLVLCVGVQOSITVDROMRYEIPRAVAFINKLDRMGAD 204  
 145 GHVDFTEVERALRVLDGAVLVLCVGVQOSITVDROMRYEIPRAVAFINKLDRMGAD 204  
 183 PARALQOM-----RSKLNHTAFMQIPMGLEGNFGIVDLIEERATVPDGF 229  
 PKVLNQVLSGAKLRHNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVVADVPNM 264  
 205 PKVLNQVLSGAKLRHNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVVADVPNM 264  
 230 SQIVYRGESIPAEALDHROELIECVANSDEQLGEMPLEEKIPISIDKLAIIRATLK 289  
 265 GQNVVADVPNMQLVMEKREELIEVSEVDQLAEAFLENDPIQANQLKAIRATVAV 324

QY 290 RSFTVFLGSAKLNKQVQPLLDVAVLEYPNPSEVQYVAILNKKDSKERTKILMNSSRN 349  
 DB 325 RKFIPIYMGSAFAFNKKGQVPLDGLDYLPCPMVEVSAL----DQNKSEKVLLAGT--P 378  
 QY 350 SHIPFVLAPFLPGRGQGLTYVRSYOGELKKGDTIYNTKTRKVRLORLARHADMMEAS 409  
 DB 379 AEPVLVALRLEGRGQLTYLRIYDGVIRKQDFINVMNGKKIKVPLVRHMSNEME-D 437  
 QY 410 TEEVYAGDICALFGIDCASGDTPTDKANSGLSMESIHPDPVISIMKPSNKNDLEKFSK 469  
 DB 438 IQEHAHQIYAVGVGVCASGDTFTD--GSVKYITMTSNANVEPWSLAVSPISDQSGQFSK 496  
 QY 470 GIGREFREDPTFKVYPTDTEKNETVIGSGEHLLEIYAQRLEREYGCPCITGKPKVAFRET 529  
 DB 497 ALNRFQKEDPTFRVGVDPESGFTIISGMGELHDIYERIRREYKVDKAVGKRVNFRET 556  
 QY 530 ITAPVFPDFTHKQSGAGQYKVIIVLEPLDEPDYTKLEFSDETGNSINPKQFVAVEX 589  
 DB 557 ITQRAEFDYLHKQSGQGGQYGVGCVIEPLPESDQKEPFDMIIIGQAIPSNFIPIAIEK 616  
 QY 590 GFLLDACERKPLSGHKLSGRLFVYLDGAAHNVDSNEISFTRAGEGALKOALANATLCLIEP 649  
 DB 617 GKREACNSSSLIGHFVBNIRIYVLTDAASHAVDSSELAFLASIIYAROCYAAARPVILIEP 676  
 QY 650 INAVEVAVANEPQGOYIAGINRRHGVITGQGVEDYFTIYADVPLNDFMGYSTELRSCTE 709  
 DB 677 VMKVELKVPTEFGVITGDMNKRKGIIVGNDQEGDDTVVCHVPLNMWGYSTALASMTQ 736  
 QY 710 GKGEYTEMYSRYQPCLPSTQEDVINKYLEATG 741  
 DB 737 GKGEFSEMYLEHNTVYSQDVQMLVNTYKASRG 768

## RESULT 9

ID Q9C641 PRELIMINARY; PRT; 754 AA.  
 AC Q9C641;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Mitochondrial elongation factor, putative.  
 GN F2G19.2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Uterback T.R.,  
 RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F2G19 genomic sequence."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBD databases.  
 DR EMBL; AC083835; AAG50635.1; -  
 DR HSSP; P13551; IELO.  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EFG\_C.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00679; EFG\_C\_1.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR PRINTS; PR00344; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATINFC.  
 DR TIGRFAMs; TIGR00484; EF-G\_1.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS00301; EFACITOR\_GTP; 1.  
 KW Elongation factor; GTP-binding; Protein biosynthesis.  
 SQ SEQUENCE 754 AA; 83178 MW; 3984E8AF9E6E977 CRC64;

Query Match

50.6%; Score 1972; DB 10; length 754;

Best Local Similarity 54.8%; Pred. No. 2,1e-128;  
 Matches 385; Conservative 118; Mismatches 187; Indels 12; Gaps 5;  
 QY 35 WSSSGVPIPEKRIKNGISAHIDSGTKTLTERLYYTGRARAKHRYKXGQVAVNDMSLT 94  
 DB 58 WKES---MDKLRNIGISAHIDSGTKTLTERLFTGRHHEHREGRGVAKRDSMDL 113  
 QY 95 ERQGITIOSAATFMKKDVNINIIDTQGVDTFEVERALAVLGVAVLVLCAGVGQCC 154  
 DB 114 EREKGIITOSAATTCWMDYKNIIIDTQGVDTFEVERALAVLGVAVLVLCAGVGQCC 173  
 QY 155 TWTVRQMKRVNVPPLTFINKLDRMGSPARALQMRSKLNTNTAFMQIPMGEGNFKGI 214  
 DB 174 SITVRQMKRVNVPPLTFINKLDRMGADPMKLNQARALKRHSAAVQPIGIESEFQGL 233  
 QY 215 VDLIERAATYFQDGSQYRVEIGIYAEALAAATDHRQELIECVANSDEGLGEMPLEKIP 274  
 DB 234 IDLHVKAFFHGGSGENVVAGDIPADMEGLVAEKRRRLIEVSEVDVLAERKFLNDEPV 293  
 QY 275 SISDLKLAIRATLRKSFPPVFLGSAKLNKQVQPLLDVAVLEYPNPSEVQYVAILNKKDD 334  
 DB 294 SASLEBEAIRKATITQTFVPVPMGSAFQKQVQPLDGVSTFLPSENVNNTAL----DQ 349  
 QY 335 SEKRTKILMNSSRHSHPVGLAPFLVGRFGQLTYVRSYOGELKKGDTIYNTKTRKVR 394  
 DB 350 NNNEBRVTLTGSPDG--PLVALAFKLEBGRFQGLTYLRYEGVIXKGDPIINVTGKRIK 407  
 QY 395 LQRLARHADMMEASTEERYAGDICALRGIDCASGDTPTDKANSGLSMESIHPDPVISI 454  
 DB 408 VRLVLRHMSNDME-DIQEHAHQIYAVFGEICASGDTFTD--GSVKYITMTSNANVEPWSL 465  
 QY 455 AMKPSNKNDLEKFSKIGREFREDPTFKYVPTENKNETVIGSGEHLLEIYAQRLEREYX 514  
 DB 466 AVQPVSKSGQGFSAALNRFOKEDPTFRGDLDPESGQITISGMGELHDIYERKRRREK 525  
 QY 515 CPCITGKPKVAFRETITAPVPPDFTHKQSGAGQYKVIIVLEPLDEPDYTKLEFSDET 574  
 DB 526 VDATGKPKRVNFRITITQABFDYLAHKQSGAGQYGVRTGVLEPLPESKKEKPEENMI 585  
 QY 575 PSSNPKQFVAVNEGFLDACRKGPLSGHKLSGRLFVYLDGAAHNVDSNEISFIRAGEBA 634  
 DB 586 VQQAIPSGFIPIAIEKGFEMANSGLIGHFVBNIRIYVLTDAASHAVDSSELAFLKAAIYA 645  
 QY 635 LKQALANATLCLIEPIMAVEVAVANEPQGOYIAGINRRHGVITGQGVEDYFTIYADVPL 694  
 DB 646 FPLCTTAARPVILIEYVMLVELKVPTEFGVITGDMNKRKGIIVGNDQEGDDTVVCHVPL 705  
 QY 695 NDMFGYSTELRSCTEGKGEYTEMYSRYQPCLPSTQEDVINKY 736  
 DB 706 NNMFGYSTLSRMTQKGSEFTMEYKHSVNSVEQAQLVNAV 747

## RESULT 10

ID Q9SHD6 PRELIMINARY; PRT; 754 AA.  
 AC Q9SHD6;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative mitochondrial translation elongation factor G.  
 GN AT2G45030.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Ronnaley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eissen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*  
 RT *thaliana*,"  
 RL Nature 402:761-768 (1999).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC007659; AD32833.1; -.  
 DR HSSP; P13551; 1ELO.  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EFG\_C.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR00795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00679; EFG\_C; 1.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATINFC.  
 DR TIGRPFAM; TIGR00484; EF-G; 1.  
 DR TIGRPFAM; TIGR00231; small\_gtp; 1.  
 DR PROSITE; PS00301; EFATOR\_GTP; 1.  
 DR Elongation factor; GTP-binding; Protein biosynthesis.  
 KW Elongation factor; GTP-binding; Protein biosynthesis.  
 SQ SEQUENCE 754 AA; 83111 MW; D3A5928C9379B7 CRC64;

Query Match 50.5%; Score 1970; DB 10; Length 754;  
 Best Local Similarity 54.8%; Pred. No. 2,8e-128;  
 Matches 385; Conservative 117; Mismatches 188; Indels 12; Gaps 5;

35 MSSSGVLPNEKINIGISAHIDSGKTLTFRVLVYGRYAKMEHVKGQGVAMDSMEL 94  
 58 WKSS---MDKLNIGISAHIDSGKTLTFRVLVYGRYAKMEHVKGQGVAMDSMEL 113  
 95 ERQGITIOSAATFTWKKVNIIDTPGVHFTIEVERALRYLDGAVLVCAVGVQCO 154  
 114 ERKGGITIOSAATFTWKKVNIIDTPGVHFTIEVERALRYLDGAVLVCAVGVQCO 173  
 155 TMTVNRQMKRYNVPFLTFINKLDRMGSNPARALQOMRSKLNHTAQMOPMGLGNPKGI 214  
 174 SITVDQMRKYEVPRVAFINKLDRMGADPWKVLNQAARAKRHSAAVQPIGLEENFOGI 233  
 215 VDLIERATYFDDDFSGIYVYGEIIPALRYAATDHOELIECVANSDEOGEWFLBKIP 274  
 234 IDLIHAKATFFHSSSGENVVAGDIPADMGLVDDKRELIEIVSEVDDVLAEXFLNDEPV 293  
 275 SISDLKLAIRATLKRSFTFVPLGSAIKKNGVQPLLDVALEVLNPNSEVONVAILNKD 334  
 294 SAELEBARIRATIAQFVVFVFGSAFKKNGVQPLLDGVVFLPSNEVNVNVA---DQ 349  
 335 SEKKTILMNSSRHNSHPVGLAFPLEVGRFGOLTVYRSYOGELKGGDTYNTRTKKYR 394  
 350 NNNEEVTLLGSPDG--PLVALAFKLEBGRFGOLTVYRVYEGVYIKGGDFINNVTGRJK 407  
 395 LQRLAHMADMASTEVEVYAGDICALFGIDCAGGPTFTKANSGLSMESIHVPDVISI 454  
 408 VPLVLRMHSNDME-DIOEHAAGQIVAVFGIECAGSOTFTD-GSVKTYTSMANPEEYMSI 465  
 455 AMKPSNKNDLEKESKIGRTREDPTFKYVFTDENKTEVYISGNGELHLIEYAQRLREYX 514  
 466 AVQPVSKDSGGFSKALNRFQKEDPTFRVGLDPESSGTIISGNGELHLIDYVERMRREX 525  
 515 CPBITGKPKVAFETITAPVPDPFTHKKQSGAGQYKIVGLEPDDPEYTLAESDET 574  
 526 VDATVKPRVNFRETTTORAEFDYLRHKQSGAGQYGRVYGVYPLPPSGKEKFEFENMI 585  
 575 FGSNIPKQFVPAVEKGFADCEKGPLSGHLSGLRFLQDGAHMDVSNISITRAGEG 634  
 586 VGOAIFSGFLPAIEKGFKAANGSLIGFVENLRIVLTDGASHAVDSBELAKRMALTYA 645

OY 635 LKQALANATLCLIEBIMAVEVAPNEFOGQVYAGINRRHGVTGGVDEYFTLYADVPL 694  
 DB 646 FLICVTAARPVILEBVMVELVKPTEFOGTAVGIDINKSGIIVGNDQGDSDSVITANVPL 705  
 OY 695 NMFVSTELRCECTGKGEYTMYSRXPCLPSTBEDYINX 736  
 DB 706 NMFVSTELRCECTGKGEYTMYSRXPCLPSTBEDYINX 747

## RESULT 11

OY 69564 PRELIMINARY; PRT; 757 AA.

AC OY 69564;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Mitochondrial elongation factor G.  
 GN MEF-G.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_Taxid=39947;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Kato A., Fujita S., Komeda Y.;  
 RT "Identification and characterization of the gene encoding the  
 RT mitochondrial elongation factor G in rice."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB040052; BAB1515.1; -.  
 DR EMBL; AB040051; BAB1515.1; -.  
 DR HSSP; P13551; 1FNM.  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EFG\_C.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR00795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00679; EFG\_C; 1.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATINFC.  
 DR TIGRPFAM; TIGR00484; EF-G; 1.  
 DR TIGRPFAM; TIGR00231; small\_gtp; 1.  
 DR PROSITE; PS00301; EFATOR\_GTP; 1.  
 KW Elongation factor; GTP-binding; Protein biosynthesis.  
 SQ SEQUENCE 757 AA; 84083 MW; 93D1156FC1955182 CRC64;

Query Match 50.4%; Score 1964; DB 10; Length 757;  
 Best Local Similarity 53.3%; Pred. No. 7,4e-128;  
 Matches 394; Conservative 117; Mismatches 214; Indels 14; Gaps 6;

OY 3 LIGAAVVALGRGAPASIGMORQVNMKACWSSGVIPNEKINIGISAHIDSGKTL 62  
 DB 31 LSAALAAASARQMPASA--LRARDEKVAWRSS---MRRNRIGISAHIDSGKTL 84  
 OY 63 TERVLVYGRYAKMEHVKGQGVAMDSMELERQGITIOSAATFTWKKVNIIDTP 122  
 DB 85 TERVLVYGRYAKMEHVKGQGVAMDSMELERQGITIOSAATFTWKKVNIIDTP 144  
 OY 123 GHVDFTEVERALRYLDGAVLVCAVGVQCOCTMTVNRQMKRYNVPFLTFINKLDRMG 182  
 DB 145 GHVDFTEVERALRYLDGAVLVCAVGVQCOCTMTVNRQMKRYNVPFLTFINKLDRMG 204  
 OY 183 PARALQOMRSKLNHTAQMOPMGLGNPKGIIVDLIERATYFDDDFSGIYVYGEIIPAL 242  
 DB 205 PAKVLNQAARAKRHSAAVQPIGLEBEEGLVDVLELAVYFEGSGGNVRLICVPSNM 264  
 OY 243 RAAATDHOELIECVANSDEOGEWFLBKIPISIDLKLAIRATLKRSFTFVPLGSAIK 302  
 DB 265 QULVMEKRELIEVSEVDDDLAEFLNDEPQANQQLAIRATYARKEFIVYMSARK 324  
 OY 303 NKGVQPLLDVALEVLNPNSEVONVAILNKDSEKKTILMNSSRHNSHPVGLAFPLEV 362





RA Yaeuda M., Tabata S.  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.",  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AF003596; BAB76037.1; -.  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EFG\_C.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF006679; EFG\_C.1.  
 DR Pfam; PF00009; GTP\_EFTU.  
 DR Pfam; PF03144; GTP\_EFTU\_D2.1.  
 DR PRINTS; PR00315; ELONGATNFC.  
 DR TIGRPFAMs; TIGR00484; EF-G.1.  
 DR TIGRPFAMs; TIGR00231; small\_GTP.  
 DR PROSITE; PS00301; EFACOR\_GTP.1.  
 KW Elongation factor; Complete proteome.  
 SQ SEQUENCE 692 AA; 76144 MW; F337ACBF03F4DD64 CRC64;

Query Match 37.8%; Score 1472; DB 16; Length 692;  
 Best Local Similarity 44.9%; Pred. No. 1e-93;  
 Matches 319; Conservative 124; Mismatches 223; Indels 44; Gaps 11;

QY 42 PNEKINIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKDGAVGAVMSMELERQGIT 101  
 DB 6 PLEKVNIGIAHIDAGKTTTERRILFYSGIHKIGEV--HEGT-AVTDMMDGERRGIT 62  
 QY 102 IGSAAFTMMKDVNINIDTPGHVDFTIEVERLRLVDGAVLVLCVGVQCCTMTVNRQ 161  
 DB 63 ITAAASTSWKDVQINIDTPGHVDFTIEVERSMRLVDGAVLFCVGVQVQPOSEVTVNRQ 122  
 QY 162 MKRYNPFLLFIKLDPMGNSPARALQOMRSKLHNHTAFMQIPMGLEGNFKGIIVDIER 221  
 DB 123 ADYKTPRIAFINKMRTGNAFVHGEQRDRLRANAIAIQLPISGENDPKGIIVDLVRK 182  
 QY 222 AIYFDGFSQIVYGEIPAEELRAATDHRQELIECVANSDEQ-----GEMPLEEKIPS 275  
 DB 183 AYYMNDQGTIDIBETIIPADLOQVESEYTKLVEAVAETDDLSKYFQDEPLTEBEINS 242  
 QY 276 ISLKAIRATLKRSFTPLFGLSALKNGVQPLDVALELYLPNBEV-----QNTA 327  
 DB 243 -----ALRGTIAGTIVPLCGSAFKNGVQMLDAVDYLPAPPEVPIQGTLLNGDA 296  
 QY 328 ILNKDQSEKTKILNNSRHNSHPFVGLAFPLEVGRFGQLTVYRSYQSELKKGDITVNT 387  
 DB 297 IERRADNE-----PLAALAEKIMADPIYGRILTVRVYSGVLKKGSIVLNA 341  
 QY 388 RTRKVRLORLARMAHADMMEASTEVEYAGDICALFGI-DCASGDTFTDKANGSLMESIH 446  
 DB 342 TKRKKERISRLVLMKADDRQ-DVEELRAGDGLAALGLKDTLTGDTITDE-GAVVLESIF 399  
 QY 447 VPPVVISIAMKPNKNDLEKFSKIGRFTREDPTFKYPTENKEVVISGMELEHLEIYA 506  
 DB 400 IEPVVISVANPEPTKNDMDLSKALQSLSEBDPTFVNVDPEPTNQVIVAGMGLHEIIV 459  
 QY 507 ORLEREYGCCTIGCKPKVAFRETTIAPVP-FDFTHKQSGAGQYKGVLEPLRPEDY 565  
 DB 460 DRMLRFKFKYANGAQVAVRETIIRKPVTVNVEGKFRQSGKQYGVHVVINLEBGEFG-- 517  
 QY 566 TKLFSDETFGSIIPKQVPAVEKFLDACEKPLSGHKLGLFVYLDGAAHMDVDSNEI 625  
 DB 518 TGEFYSKIVGVVPREYIGPAQGMKSCESGILMGVPLIDVKATLVHGSYHVDVSSM 577  
 QY 626 SFPRABEGALKALANATLCILEPIAVENVVAVNEQOYIAGINRRHGVITGQDVEDY 685  
 DB 578 AAFIASMALKKAVLVLAASPLLEPMKVEVEVEDYIGVIGDLISRRQGISQSTEGQI 637  
 QY 686 FTLYADVPLNDMGYSTELRSCTEGGEYTMESRYOPLCPSQEDVINK 735  
 DB 638 AKVASVPLATMGVATIDIRSKTQGRGIFTMERSHYEBVPRVVAETIIAK 687

## RESULT 14

QY 09RNG0 PRELIMINARY; PRT: 682 AA.  
 ID 09RNG0  
 AC 09RNG0  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Elongation factor G (Fragment).  
 GN EFG.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OC NCBI\_Taxid=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RU-1;  
 RA Kolesnikow T., Gekas S., Lee A.;  
 RT "Identification of Helicobacter pylori Antisense";  
 RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF181877; AAF04270.1; -.  
 DR HSSP; P13551; IDAR.  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EFG\_C.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF006679; EFG\_C.1.  
 DR Pfam; PF00009; GTP\_EFTU.1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2.1.  
 DR PRINTS; PR00315; ELONGATNFC.  
 DR TIGRPFAMs; TIGR00484; EF-G.1.  
 DR TIGRPFAMs; TIGR00231; small\_GTP.1.  
 DR PROSITE; PS00301; EFACOR\_GTP.1.  
 KW GTP-binding.  
 FT NON TER 1  
 SQ SEQUENCE 682 AA; 75939 MW; FC8610CC97CB584E CRC64;

Query Match 37.7%; Score 1470; DB 2; Length 682;  
 Best Local Similarity 45.5%; Pred. No. 1.4e-93;  
 Matches 315; Conservative 129; Mismatches 231; Indels 18; Gaps 11;

QY 47 NIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKDGAVGAVMSMELERQGIT 106  
 DB 1 NIGISAHIDAGKTTTERRILFYTGSHKIGEV--HDG-AATMDMEQEKERGIT 57  
 QY 107 TPTMKDVNINIDTPGHVDFTIEVERLRLVDGAVLVLCVGVQCCTMTVNRQMKRN 166  
 DB 58 TTCFMDHQINIDTPGHVDFTIEVERSMRLVDGAVSVFCVGVQVQPOSEVTVNRQANKYG 117  
 QY 167 VPFLLFIKLDPMGNSPARALQOMRSKLHNHTAFMQIPMGLEGNFKGIIVDIERAIYFD 226  
 DB 118 VRIYFVNMKMDIGANFVNVENQIKLRKANVPINIGADPTIIGVIDLVQMAIWN 177  
 QY 227 GDFSOIVRGG-EIPAEELRAATDHRQELIECVANSDEQLEGMLEBKIPISIDUKLAIR 284  
 DB 178 NE-TWGAQYDVEISDLEKAKQYREKLVEVAQBDALEMEKYLGSELENIIEIKGK 236  
 QY 285 RATLRSFTPVPLGLSALKNGVQPLDVALELYLPNBEVQVNAIILNKDQSEKTKILMN 344  
 DB 237 TGLNMSLVPMICGSSFNKRGVOTLLDAVIDLPAPEVVDI---RGIDPTEEEVFK 292  
 QY 345 SSRHNSHPFVGLAFPLEVGRF-GQLTVYRSYQSELKKGDITVNTTRKKVRLQRLARMA 403  
 DB 293 SS--DDGEFAGLAFKIMDTPFGQLTFPARYVRGKLSSSYVNSYTKDKKERGRLLRKH 350  
 QY 404 DMEASTEVEYAGDICALFGI-DCASGDTFTDKANGSLMESIHVDPVISIAMKPNKN 462  
 DB 351 NKRE-DIKEVYAGECIAFGLKDTLTGDTLCEKNA-VLEBMEPEPIHIAVEPTKA 408  
 QY 463 DLEKSKIGRTRRDPPTFKYPTENKEVVISGMELEHLEIYAQRLEREYGCCTIGKP 522  
 DB 409 DQKMGVALGKLAEDPSFRVNTQETGOTLIGGMELEHLEIIVDLRKRFEVVAEIQGP 468

```

Qy 523 KVAFRTTAPVPPDTHKKSGAGQYGVIGVLEPDLFEDYTKLEFSDETFGSNIPKQ 582
Db 469 QVAFETTTSSVSKHKYAKSGGRGQVHPFKLEPKKEFG--SGEYFNEISGGVIPK 526
Qy 583 PVPAYEKGLDACEKGPLSGHLSGLRFLADGAAHHVDSNESPFRAGEGALKOALANA 642
Db 527 YLPADVKGQEMQNVNLGVYFVDPKVTLLYDSYHVDVSEMAFKIAGSMFKKASRA 586
Qy 643 TLTILEPIPAVEVVAENEFQGVYAGINRRHGVITGQDGVEDYFTLYADVPILNMFYST 702
Db 587 NVVLEPMMKVEVEVEEYMGVIGDLNRRGOINMDHLGIKYNVAFVPLEVMFGYST 646
Qy 703 ELRSCTEGGEYTMESRYOPLPSTQEDVINK 735
Db 647 DLRSATQGRGYSMEDHYGEVPSNISKIIVEK 679

RESULT 15
Q9AF78 PRELIMINARY; PRT; 700 AA.
ID Q9AF78;
AC 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Elongation factor G.
GN FUS.
OS Arthrobacter sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococccaceae; Arthrobacter.
OX NCBI_TaxId=1667;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=515;
RA Berchet V., Russell N.J., Normand P., Gounot A.M.;
RT "Cloning and sequencing of the fus gene coding for elongation factor G
RL from the arctic psychrotroph Arthrobacter sp. strain 515.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238083; CAC36321.1; -.
DR HSSP; P13551; 1DAR.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EF-G.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_BFTU; 1.
DR PRINTS; PR00315; ELONGATINFT.
DR TIGRFAM6; TIGR00484; EF-G; 1.
DR TIGRFAM6; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW GTP-binding.
SQ SEQUENCE 700 AA; 77478 MW; 1CC32B349330AA61 CRC64;

Query Match 36.6%; Score 1427.5; DB 2; Length 700;
Best Local Similarity 43.2%; Pred. No. 1,3e-90;
Matches 304; Conservative 134; Mismatches 240; Indels 25; Gaps 10;

Qy 45 KIRNIGISAHISGKTLTERVLYYTG--RIAKMEHVKGKQGVAVMDSMELERQGIT 101
Db 7 KVRNIGIMAHIDAGKTTTTERILFYGVNHHKIGETH---DG-ASTTDMQEKERGIT 60

Qy 102 IGSAAFTFMKDVNINIIDTGHVDFTIEVERALRVLDGAVLVLCVAGVQCOTMTVNRQ 161
Db 61 ITSAAVTCFEMNQNIINIIDTGHVDFTIEVERSLRVLDGAVAVFDGKEGVEPQSETVMRQ 120

Qy 162 MKRYNVPELTFLINKLDRMGSNPARALQMRSKLNHTAFMQIPMGLGNPKGIVDLIEER 221
Db 121 ADKYNVPRIKCFVKNQMDKADAFYFVDTIISRLGVKPLVMQPLIGAENDFIGVDDLVMR 180

Qy 222 AIYFDGDFSOIYRYG-----EIPALRLAATDHRQELIECVANSDEQLEGMFLEKIPS 275
Db 181 ALVWPDSDSKGVDYMGAKYEIOETPADLKEABEYRALVETVAESSEELMEKYLEGEIT 240

```

```

Qy 276 ISDLKAIIRATLKESFTPEVFLGSAKKNKQVOLLDAVLBYLNPSEVQNAIILKKDSD 335
Db 241 EDELKAG;RKMTINSELVDFVFCGSAFKNRGVQPMLDVAVDVLPNPLDVPMTGHPDRBE 300
Qy 336 KEKTIILNNSHHNSHPVGLAFPLEVGR-FQGLTYVASYQGLKKQDTIYTRTRKKYR 394
Db 301 KELTR-----KPSSEPPSALAFLKATHPFFQGLTFIHYSGHVAAGQVAVNSTGKKER 355
Qy 395 LQRLARHMDMWEASTEVEYVAGDICALFGI--DCAGDPFTDKNAGSLMESIHVDDPVIS 453
Db 356 IGLRQMHAN-KEMVVDGATGHYIAIGLKDTTGGTLCBPANO-IYLSHSPPEPVIS 413
Qy 454 IAKPSNKNLDEKFSKGIGRFTREDPTKVPFDTENKSTVLSGMGLHLIETVQRLEREY 513
Db 414 VAIEPNTGDDQEKSTAIQKLSAEDPTQVSLNEDTGTITAGMGLHLHDLIVDMRRERF 473
Qy 514 GCPCTIGKPKVAFRETTIAPVP-FPETHKKSGAGQYGVIGVLEPDLFEDYTKLEFSD 572
Db 474 KVEANVGRQVAVRETIKRAVERHDYTHKKQYSGSGQPAKIQIAIEPLDTSEGLYERPN 533
Qy 573 EFTGSNIPKQFVPAYEKGPLDACEKGPLSGHLSGLRFLADGAAHHVDSNESPFRAGE 632
Db 534 KVTGRIRPREYIPSVDAQIQDRLNDGVLAGIPVGIKATLIDGAYHVDSSMARKLRGR 593
Qy 633 GALKQALANATLCILEPIPAVEVVAENEFQGVYAGINRRHGVITGQDGVEDYFTLYADY 692
Db 594 MAFKRAARKANPILLEPLMDVETRPEEYMGVIGDLNRRGOINMDHLGIKYNVAFVPLEVM 653
Qy 693 PLNMFSGISTELRSCTEGGEYTMESRYOPLPSTQEDVINK 735
Db 654 PLSGMFGYIGDLRSKTOGRAVAVSMTHSYAEVFKAFADBIQK 696

```

Search completed: July 14, 2003, 18:21:01  
 Job time : 39.8872 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 7.82169 Seconds  
(without alignments)  
3987.653 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898  
Sequence: 1 MRLGAAVAALGRGRAPAS.....INKYLATGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3343	83.2	752	1 EFGM_RAT	Q07803 Rattus norv
2	1987	51.0	761	1 EFGI_YEAST	P25039 Saccharomyc
3	1765	45.3	695	1 EFGI_TREPA	O83748 treponema p
4	1762	45.2	693	1 EFGI_BORBU	O30913 borrelia bu
5	1493	38.3	694	1 EFGI_SYMP6	P18667 synechococc
6	1489	38.2	691	1 EFGI_HELPU	O92424 helicobacte
7	1487	38.1	691	1 EFGI_HELPU	P56002 helicobacte
8	1462	37.5	702	1 EFGI_YERPE	O824b3 yersinia pe
9	1460.5	37.5	691	1 EFGI_CAMJE	O93116 campylobact
10	1454	37.3	699	1 EFGI_SPIPL	P13550 spirulina p
11	1454	37.3	699	1 EFGI_HABIN	P43925 haemophilus
12	1453	37.3	700	1 EFGI_RALSO	O84xm7 ralstonia s
13	1438.5	36.9	700	1 EFGI_FASMU	P57338 pasteurella
14	1438	36.9	702	1 EFGI_BUCAI	P57593 buchnera ap
15	1434	36.8	703	1 EFGI_SALTY	P26229 salmoneilla
16	1431.5	36.7	691	1 EFGI_THETH	P13551 thermus the
17	1430	36.7	701	1 EFGI_MYCTU	O53790 mycobacteri
18	1428.5	36.6	696	1 EFGI_RHTLO	O98459 rhizobium l
19	1427	36.6	703	1 EFGI_ECOLI	P02396 escherichia
20	1426.5	36.6	691	1 EFGI_SYNY3	P74428 synchocyst
21	1422	36.5	701	1 EFGI_MYCLE	P30767 mycobacteri
22	1419	36.4	703	1 EFGI_RALSO	O84x10 ralstonia s
23	1416.5	36.3	701	1 EFGI_NELMA	O93407 neisseria m
24	1414.5	36.3	694	1 EFGI_BRUME	O84xnp3 brucella me
25	1413.5	36.3	701	1 EFGI_NELMA	O93418 neisseria m
26	1408.5	36.1	705	1 EFGI_XYLFA	O93490 xyella fa
27	1407.5	36.1	692	1 EFGI_THEMA	O91444 pseudomonas
28	1406	36.1	702	1 EFGI_PSEAR	O91444 pseudomonas
29	1401	35.9	687	1 EFGI_CLOAB	P75544 clostridium
30	1399.5	35.9	688	1 EFGI_MYCPN	P75544 mycoplasma
31	1398.5	35.9	699	1 EFGI_AQUAE	O66428 aquilex aeo
32	1398	35.9	699	1 EFGI_RICCN	O92973 rickettsia
33	1397	35.8	708	1 EFGI_STRCO	P40173 streptomyce

34	1396	35.8	688	1 EFGI_CLOPE	Q6xhei clostridium
35	1395	35.8	695	1 EFGI_SYNY3	P28371 synchocyst
36	1395	35.8	698	1 EFGI_VIBCH	O93427 vibrio chol
37	1392.5	35.7	691	1 EFGI_BACSU	P80868 bacillus su
38	1388	35.6	695	1 EFGI_LISIN	O92715 listeria in
39	1388	35.6	699	1 EFGI_RHIME	O924h2 rhizobium in
40	1387	35.6	695	1 EFGI_LISMO	O84x21 listeria mo
41	1387	35.6	699	1 EFGI_RICPR	P41084 rickettsia
42	1384.5	35.5	706	1 EFGI_PSEAR	O91444 pseudomonas
43	1380	35.4	692	1 EFGI_MYCPN	O93428 mycoplasma
44	1373	35.2	701	1 EFGI_MYCTU	P09852 micrococcus
45	1372	35.2	699	1 EFGI_AGRTU	P70782 agrobacteri

## ALIGNMENTS

RESULT 1	ID	EFGM_RAT	STANDARD;	PRT;	752 AA.
AC	Q07803;				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	Elongation factor G, mitochondrial precursor (MEF-G).				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93324327; PubMed=8332461;				
RA	Barber C.S., Makris A., Patriotic C., Bear S.B., Tsichlis P.N.;				
RT	"Identification of the gene encoding the mitochondrial elongation factor G in mammals."				
RT	factor G in mammals."				
RL	Nucleic Acids Res. 21:2641-2647(1993).				
CC	- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLATION				
CC	OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE				
CC	RIBOSOME.				
CC	- SUBCELLULAR LOCATION: Mitochondrial.				
CC	- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES WITH THE HIGHEST LEVEL				
CC	IN LIVER, THYMUS AND BRAIN.				
CC	- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.				
CC	EF-G/EF-2 SUBFAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; L14684; AAA1107.1; -.				
DR	PIR; S40780; S40780.				
DR	HSSP; P13551; IPNM.				
DR	InterPro; IPR004540; EF-G.				
DR	InterPro; IPR000640; EF-G.				
DR	InterPro; IPR004161; EFTU_D2.				
DR	InterPro; IPR000795; EF_GTPbind.				
DR	InterPro; IPR005225; Small GTP.				
DR	Pfam; PF000079; GTP_EFTU; 1.				
DR	Pfam; PF00679; EFG_C; 1.				
DR	Pfam; PF03444; GTP_EFTU_D2; 1.				
DR	TIGRFAMs; TIGR00231; small_GTP; 1.				
DR	TIGRFAMs; TIGR00484; EF-G; 1.				
DR	PROSITE; PS00301; EFACITOR_GTP; 1.				
DR	Elongation factor; Protein biosynthesis; Mitochondrion;				
KW	Transit peptide; GTP-binding.				
FT	TRANSIT 1 ?				
FT	CHAIN 1 ?				
FT	NP_BIND 54 61				
FT	NP_BIND 121 125				
FT	MITOCHONDRION (POTENTIAL).				
FT	ELONGATION FACTOR G.				
FT	GTP (BY SIMILARITY).				
FT	GTP (BY SIMILARITY).				

FT NP\_BIND 175 178 GTP (BY SIMILARITY).  
 SQ SEQUENCE 752 AA; 83770 MW; DFB6108D38A72E4 CRC64;  
 Query Match 83.2%; Score 3243; DB 1; Length 752;  
 Best Local Similarity 84.1%; Pred. No. 1.7e-206;  
 Matches 636; Conservative 37; Mismatches 75; Indels 8; Gaps 5;

QY 1 MRLGAAVAALGRG---RAPASLGWQRKQVNMKACRWSSSGVLPNEKIRNIGISAHIDS 57  
 DB 1 MRL--RITAGIGRGLPRVPALIGWQGANWKTRWSSSGSIPNEKIRNIGISAHIDS 58  
 QY 58 GKTLTTERLYYGRGAKMHEVKGKGVAVNDMSLELROGRTTISATFTMKGVNIN 117  
 DB 59 GKTLTTERLYYGRGAKMHEVKGKGVAVNDMSLELROGRTTISATFTMKGVNIN 118  
 QY 118 IIDTPGHVPTIEVERALRLDGAVALVCAVGVOGQMTVNMQRVNVPLFTINKD 177  
 DB 119 IIDTPGHVPTIEVERALRLDGAVALVCAVGVOGQMTVNMQRVNVPLFTINKD 178  
 QY 178 RMGNSPARALQOMRSKLNHTAFMOIPMGLEGNFKGIVDLIBERATYFDGDFSOIVRYGE 237  
 DB 179 RMGNSPARALQOMRSKLNHTAFMOIPMGLEGNFKGIVDLIBERATYFDGDFSOIVRYGE 238  
 QY 238 IPMBLBAATDNRQSLIECVANSDEQLGEMFLEBKIPISDLKLAIRRALTKRSFTPVTL 297  
 DB 239 IPADLBAADNRPELIECVANSDEQLGEMFLEBKIPISVDLKLAIIRRALTKRSFTPVTL 298  
 QY 298 GSALKNKGVOPLDAVLEVLNPSEVONAYILNKDSEKTKILNNSRNHNPVGLA 357  
 DB 299 GSALKNKGVOPLDAVLEVLNPSEVONAYILNKDSEKTKILNNSRNHNPVGLA 357  
 QY 358 FPLEVGRFGQLYVRSYOGELKKGDITNTTRTKKVALRQLARMAHADMMASTEEVYAD 417  
 DB 358 FPLEVGRFGQLYVRSYOGELKKGDITNTTRTKKVALRQLARMAHADMMASTEEVYAD 416  
 QY 418 ICLFELIDASGDTFTDKANSGLSMESINHPDVISIANKPSKNDLEKPSKIGFTPE 477  
 DB 417 ICLFELIDASGDTFTDKANSGLSMESINHPDVISIANKPSKNDLEKPSKIGFTPE 476  
 QY 478 DPTFKYFPTENETVYISGMGELHLEIYAORLEREGCPCITGPKVAPRETITAVPPD 537  
 DB 477 DPTFKYFPTENETVYISGMGELHLEIYAORLEREGCPCITGPKVAPRETITAVPPD 536  
 QY 538 FTHKKSOGGA-GQYKVGIVLEPLDEDTYKLEFSDTEGNSIPKQFPAVEKGFPLDACE 596  
 DB 537 ITHIKSSRVVPASMGKIVGLPELPEDELPKLEFSDTEGNAVPKQFPAVEKGFPLDACE 596  
 QY 597 KGPLSGHKLSGLRFPVODGAHNVDSNEISFIRAGGALKQALASATLCIIEPIMSVEVY 656  
 DB 597 KGPLSGHKLSGLRFPVODGAHNVDSNEISFIRAGGALKQALASATLCIIEPIMSVEVY 656  
 QY 657 APPEFGOVYAGINRRHGVITGQDVEDFTLYADVPLNDMPGYSTELSCTEGKEGYTM 716  
 DB 657 APPEFGOVYAGINRRHGVITGQDVEDFTLYADVPLNDMPGYSTELSCTEGKEGYTM 716  
 QY 717 EYSRYOPLPSTQEDVINKLEATGQLPVKGRKAKN 752  
 DB 717 EYSRYOPLPSTQEDVINKLEATGQLPVKGRKAKN 752  
 RESULT 2  
 EFGI\_YEAST STANDARD; PRT; 761 AA.  
 AC P25039; Q99360; 01-MAY-1992 (Rel. 22, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Elongation factor G 1, mitochondrial precursor (MEF-G-1).  
 GN MEF1 OR YLR069C OR L2195.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;  
 NCBI\_TaxId=4932;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92037620; PubMed=1935960;  
 RA Vamvutae A., Ackeman S.H., Tzagoloff A.;  
 RT "Mitochondrial translational-initiation and elongation factors in  
 Saccharomyces cerevisiae";  
 RL Eur. J. Biochem. 201;643-652(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Andre B., Urrestarazu L.A.;  
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION  
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 CC RIBOSOME.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X58378; CAA41267.1; -;  
 CC EMBL; X94607; CAA64315.1; -;  
 CC EMBL; Z73241; CAA97626.1; -;  
 CC PIR; S17025; S17025.  
 CC PIR; S20179; S20179.  
 CC HSSP; P13551; 1FNM.  
 CC SGG; S0004059; MEF1.  
 CC InterPro: IPR004540; EF-G.  
 CC InterPro: IPR000640; EFG C.  
 CC InterPro: IPR004161; EFTU D2.  
 CC InterPro: IPR000795; EF\_GTPbind.  
 CC InterPro: IPR005225; Sma11 GTP.  
 CC Pfam; PF00009; GTP\_EFTU; 1.  
 CC Pfam; PF00679; EFG\_C; 1.  
 CC Pfam; PF03144; GTP\_EFTU D2; 1.  
 CC PRINTS; PR00315; ELONGATNFACT.  
 CC TIGRfam; TIGR00231; sma11 GTP; 1.  
 CC TIGRfam; TIGR00484; EF-G; 1.  
 CC PROSITE; PS00301; EFACITOR GTP; 1.  
 CC Elongation factor; Protein biosynthesis; Mitochondrion;  
 CC Transist peptide; GTP-binding.  
 CC TRANSIT 1 ?  
 CC CHAIN ? 761  
 CC NP\_BIND ? 84  
 CC NP\_BIND 148 152  
 CC NP\_BIND 202 205  
 CC CONFLICT 66 66  
 CC CONFLICT 233 233  
 CC CONFLICT 478 478  
 CC CONFLICT 629 629  
 CC SEQUENCE 761 AA; 84573 MW; 3E2C534509B09103 CRC64;  
 Query Match 51.0%; Score 1987; DB 1; Length 761;  
 Best Local Similarity 56.2%; Pred. No. 1.4e-123;  
 Matches 392; Conservative 117; Mismatches 172; Indels 16; Gaps 8;

QY 45 KIRNIGISAHIDSGKTLTTERLYYGRGAKMHEVKGKGVAVNDMSLELROGRTTIS 104  
 DB 69 KIRNIGISAHIDSGKTLTTERLYYGRGAKMHEVKGKGVAVNDMSLELROGRTTIS 128  
 QY 105 AATFTMW---KDVNINIIDTPGHVPTIEVERALRLDGAVALVCAVGVOGQMTVNR 160  
 DB 129 AATFTMWKEGKGVHFNIDTPGHVPTIEVERALRLDGAVALVCAVGVOGQMTVNR 188  
 QY 161 QMKRNVNPLFTINCLDMGNSPARALQOMSKLNHTAFMOIPMGLEGNFKGIYDLIEE 220  
 DB 189 QMKRNVNPLFTINCLDMGNSPARALQOMSKLNHTAFMOIPMGLEGNFKGIYDLIEE 248



```

Db      595 VVLEPIKRVSEVGEPOFQSGVFGILNQRGVVSSADDEQFSRVDALEVPSEKFGFSTVL 654
QY      705 RSTCEKGETMYEYSRQPCLPSTQEDVINKYLE 738
Db      655 RSTQGAKEYSMERAKGAPQGVTSLLIKYDE 688

RESULT 4
EFG1_BORBU STANDARD; PRT; 693 AA.
AC      030913, 051490,
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Elongation factor G 1 (EF-G 1).
GN      FUSA OR FUS1 OR BB0540.
OS      Borrelia burgdorferi (Lyme disease spirochete).
OC      Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX      NCBI_TaxId=139;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=DK1;
RA      Sonderbye L., Hindersson P., Pedersen S.;
RL      Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35210 / B31;
RA      MEDLINE=98065943; Pubmed=9403685;
RA      Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA      Lathigra R., White O., Ketchum K.A., Dodson R., Hickey B.K., Gwinn M.,
RA      Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA      Peterson J., Karlyavega A.R., Quackenbush J., Salzberg S., Hanson M.,
RA      van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA      Utecherback T., Wachtel N., McDonald L., Artlich P., Bowman C.,
RA      Garland S., Fujii C., Cotton M.D., Horek K., Roberts K., Hatch B.,
RA      Smith H.O., Venter J.C.;
RT      "Genomic sequence of a Lyme disease spirochete, Borrelia
RT      burgdorferi."
RL      Nature 390:580-586(1997).
CC      -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLATION
CC      OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC      RIBOSOME.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC      EF-G/EF-2 SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC      or send an email to license@1sb-sib.ch).
CC      -----
DR      EMBL; AF021260; AAB71893.1; -.
DR      HSSP; A5E01155; AAC68897.1; -.
DR      HSSP; P13551; 1ELO.
DR      TIGR; BB0540; -.
DR      InterPro; IPR004540; EF-G.
DR      InterPro; IPR000640; EFG_C.
DR      InterPro; IPR004161; EFTU_D2.
DR      InterPro; IPR000795; EF_GTPbind.
DR      InterPro; IPR005225; Small_GTP.
DR      Pfam; PF00009; GTP_EFTU; 1.
DR      Pfam; PF00679; EFG_C; 1.
DR      Pfam; PF03144; GTP_EFTU_D2; 1.
DR      TIGRfam; TIGR00231; small_GTP; 1.
DR      TIGRfam; TIGR00484; EF-G_1.
DR      PROSITE; PS00301; EFATOR_GTP; 1.
KW      Elongation factor; Protein biosynthesis; GTP-binding;
KW      Complete proteome.
FT      NP_BIND 13 20 GTP (BY SIMILARITY).

```

```

FT      NP_BIND 80 84 GTP (BY SIMILARITY).
FT      NP_BIND 134 137 GTP (BY SIMILARITY).
FT      CONFLICT 201 201 D -> E (IN REF. 1).
FT      CONFLICT 208 208 S -> N (IN REF. 1).
FT      CONFLICT 212 212 I -> M (IN REF. 1).
FT      CONFLICT 216 216 T -> A (IN REF. 1).
FT      CONFLICT 236 236 T -> I (IN REF. 1).
FT      CONFLICT 243 243 T -> I (IN REF. 1).
FT      CONFLICT 243 243 E -> S (IN REF. 1).
FT      CONFLICT 304 304 I -> T (IN REF. 1).
FT      CONFLICT 331 331 I -> G (IN REF. 1).
FT      CONFLICT 484 484 R -> A (IN REF. 1).
FT      CONFLICT 504 504 D -> A (IN REF. 1).
FT      CONFLICT 513 513 R -> K (IN REF. 1).
FT      CONFLICT 548 548 R -> K (IN REF. 1).
FT      CONFLICT 671 671 T -> A (IN REF. 1).
FT      CONFLICT 678 678 D -> R (IN REF. 1).
FT      CONFLICT 689 689 D -> R (IN REF. 1).
SQ      SEQUENCE 693 AA; 77572 MW; F8A44C38D58F9B1E CRC64;

Query Match 45.2%; Score 1762; DB 1; Length 693;
Best Local Similarity 51.8%; Pred. No. 8,9e-109;
Matches 361; Conservative 126; Mismatches 186; Indels 22; Gaps 10;

QY      45 KIRNIGISAHIDSGKTTLTERTVLYTGRJAKNHEVKGXGQGVANVDSMELERQGITIOS 104
Db      5 KLRNIGISAHIDSGKTTLTERTLIFCNKHAHVEYKXGQGVATWDSMELEREGITIAS 64
QY      105 AATFTMKDVNINIIDTDEGHVDFTIEVERALRVLDGAVLYLCVAGVQCQMTNVRQMR 164
Db      65 AATHEWMDQFINIIDTPEHVDFTIEVERSLNVDGAILVDSVAGVQSQSTIVRQLKR 124
QY      165 YVVPFLTFINKDKRMGNPARALQOMRSKLNNTAFMOIPMGLNFGKIVDLIEBAIY 224
Db      125 YVPLPLAFNKKDKKGNAPYNNKQDLRSKLDNSVLMQPILEDKHGIVDLVLMKAY 184
QY      225 FPG-DPSQIVRYGEIIPAEIRAAATDRQELIECVANSBQGLCEMFLKXIPISDLKAI 283
Db      185 FPGKDTETIE-KETPSDLIEAKSKREIMDTLADFDELMEHMEKEVETETIYVNT 243
QY      284 RRAATLKRSPTVFLSALKNKGVQLDAVLEYLNPSVQVYALINKDSEKETKILM 343
Db      244 RTGTALKLKCPVMSAYKNKGVQLDAVTRFLSPHDKNTALDNNNEKIDKI-- 301
QY      344 NSGRNHSHPFVGLAPPLEVGRFQLTYYRSYQGELEKGGDTYNTTRKKVRLQRLARMA 403
Db      302 ----DNEPFTVALAKLEGGQGLTYRIYGIILAKGOELNSTRSKKFKRLLIRMA 357
QY      404 DMEASTSEEVY--AGDICALFGIDCASGDTFTDYANSGLSNEIHPDVISIAMKPSN 460
Db      358 N---NTEDIEFGSGDIALFGICASGDTFCD-PSINYSWTSMFIPDPVISLSVKPYD 412
QY      461 KNDLEKFSKIGIRFREDPTFPVYPTENKEVIVSGMEHLEIYAQRLEREGPCRTG 520
Db      413 KKSADNMMAALGRFKEDPTFTYVDISNETIIGKMEHLLEVYIERKRFKAEVETG 472
QY      521 KPVAFRETTAPVFPDFTHKKQSGAGQYKIVGLPELPDEBDYTKLEFSDETGSNP 580
Db      473 MPQVAYRETTAKAFNTTHKKQSGAGQFGVAFPMERLDEGET-YEFVVLINGVYIP 531
QY      581 KQFVAVEKGFLLDACERKPLSGHKLSGLRVYLDGAHHNVDSNEISFIRAGEALKQALA 640
Db      532 TEYIPSCDQGFQKAMERGLTIGFPIVDIKITINDQYHIVDSSDIAFOLAIGAFREAVE 591
QY      641 NATLCILPEYMAVEVYANPEPQGVYIAGINRHGVITG--QGVADYFTLYADVLDNMF 698
Db      592 KAKPTLLEPIMKVTLEGPTEFGNMFGILNORGIITSLDEG--SFSKVEAEVPLESEMF 649
QY      699 GYSTELEKSTCEKGETMYEYSRQPCLPSTQEDVINK 735
Db      650 GFSTVLRSTQGAKEYSMERAKGAPQGVTSLLIKYDE 685

RESULT 5

```



```

EFG_SYNP6
ID EFG_SYNP6 STANDARD; PRT; 694 AA.
AC P18667;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUS1 OR FUS.
OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
NC NCBI_TaxID=1139;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89281486; PubMed=2499762;
RA Meng B.-Y., Shinozaki K., Sugita M.;
RT "Genes for the ribosomal proteins S12 and S7 and elongation factors
RT EF-G and EF-Tu of the cyanobacterium, Anacystis nidulans: structural
RT homology between 16S rRNA and S7 mRNA."
RL Mol. Genet. 216:25-30(1989).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X17442; CAA35495.1; -
DR PIR: S04429; S04429.
DR HSP: P1351; 2EEG.
DR InterPro: IPR0004540; EF-G.
DR InterPro: IPR000640; EF-G.
DR InterPro: IPR004161; EFTU D2.
DR InterPro: IPR000785; EF-GTPbind.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00069; GTP_EFTU; 1.
DR Pfam: PF03144; GTP_EFTU D2; 1.
DR TIGRfam: TIGR00231; small_GTP; 1.
DR TIGRfam: TIGR00484; EF-G; 1.
DR PROSITE: PS00301; EF-CTOR-GTP; 1.
DR Elongation factor; Protein biosynthesis; GTP-binding.
KW NP BIND
FT NP BIND 17 24 GTP (BY SIMILARITY).
FT NP BIND 81 85 GTP (BY SIMILARITY).
FT NP BIND 135 138 GTP (BY SIMILARITY).
SQ SEQUENCE 694 AA; 75909 MW; D1ADFCD515C95FF CRC64;

Query Match 38.3%; Score 1493; DB 1; Length 694;
Best Local Similarity 44.5%; Pred. No. 5, 1e-91;
Matches 315; Conservative 128; Mismatches 229; Indels 36; Gaps 9;

```

```

QY 281 LAIRATLKRS---FTPVFLGSLAKKKGVPLDAVLEYLPNPSHQ-----NYAI 328
DB 242 AGLRKGVLIIQGNDRVLVPLMCSAFKNGVQLLDNAVLELPSPDIPPIQGLPGGVALL 301
QY 329 LNKKDSKYEKTKILMNSSRHNSHPVGLAFLPLEVGRFQGLTVRSYGGLKKGGTYNTR 388
DB 302 RPSDEA-----PSSALAFKIMADPGRLEFVRVYSGILQKGSYYVYNT 345
QY 369 TRKRYRLQRLAMHADMMEASTEEVYAGDICALPGI -PCASGDTPTDANKSLMESIHV 447
DB 346 KKKKRVSRLLIILKAD-DRIEVDLRAGDLGAVLGLKPTGDTLCDQNP-ILLESLEPI 403
QY 448 PDPVISIMKPSNKNLDEKFSKIGRFTREDPTFKVYEDTEKTEKYSIGMGLHLEIYQ 507
DB 404 PEPVSAVNEPKTKNDMEKLSALQALSEBDPTFVSVDSERNQVINGMGLHLEIYVD 463
QY 508 RLREYGCPCITGKERVAFRETTAPVPDPPTHKKQSGAGQYGVIVGLEPLDDEYTK 567
DB 464 RMLREYKVEANIGAPQVAVRETVRAVYAEKGFVRSQSGKGQGVHVIIELEPAEPG--TG 521
QY 568 LEFSDETGSNIPKQFVAVKEGFLDACEKGPLSGHKLGLARFVLDGAMHVNDSNISF 627
DB 522 FEFVSKIVGTVPKSYVGPAGQMKETCSGLVAGYPLIDIKATLVDSYHDVDSSEMAF 581
QY 628 IPAGEALQALANATLILPEIMAVEVAPNEPQGVYAGINRRHGVITGQGVEDYPT 687
DB 582 KLAGSMALKEAVRKADPVLEBVMVEVEVPEDFISVMGNLISRSQIEGATNIGRT 641
QY 668 LVADVPVLDMPFGYSTELRSCTEGKGYTMEVRSYRQPCSTQEDYINK 735
DB 642 VSAKVPFLAMFQYATDLMSMTGSGRIPTMEFSGYEVPRNVAETIIAK 689

```

```

RESULT 6
EFG_HELP
ID EFG_HELP STANDARD; PRT; 691 AA.
AC Q9ZK24;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUS1 OR JHP1118.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Møllr D.T.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RT Nature 397:176-180(1999).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```



```

QY 42 PNEKIRNIGISAHIDSGKTTLTERVLYYTGRIAKNEHVKGDVGAVMDSMELEKRGIT 101
DB 5 PLNRIRNIGISAHIDAGKTTTERRILFYGVSHKIGEV--HDG-AATMDMMEQEKERGIT 61
QY 102 IQSAATFTFMKDVNINIIDTPGHVDFTTEVERARLVLDGAVLVLCVGVGVOQOTMVRNQ 161
DB 62 ITSAAITTCWKHQJNLIDTPGHVDFTTEVERARLVLDGAVLVLCVGVGVOQOTMVRNQ 121
QY 162 MKRYVVPFLFTINKLDRMGSNPAPALQOMRSKLNHTAFMOJPMGLEGNFKGIVDLIER 221
DB 122 ANKGVPRIVFNKNDRIKANPYNVNOIKLILKANPVEKINPIGHEDFIIGVLDVQMK 181
QY 222 AIYFGDSFOIVRYG--EIPAEIRAAATDHRQELIECVANSDEQLGEMLEKIPISIDL 279
DB 182 AIVMNNNE-TWGAKEYDVEIIPDLLEKAKEREKLVAVAEQBALMEKYLGEELISIERI 240
QY 280 KLAIRATIKRSTPTPLGSLAKNGVOPLDVALEYLPNPEVQNYALINKKDSKERT 339
DB 241 KKGIRAGCLNMSLVPLCGSSFFKNKGVOTLLDAVIDLYAPTEVVDI---KGIDPKTER 296
QY 340 KILMNSRHNHPVGLAPFLEVGRF-GQLTYVRSYQGEIKKGDITYNTRTKKVALRL 398
DB 297 EYFVAVSS--DDEPFLAKIMTDPVQGLTFRVYRGKLESGSYVNSTKXKERVNRL 354
QY 399 ARWADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHVPDVISIAWK 457
DB 355 LKMHNSKRE-DIKEYVYAGEICAFVGLKDTLGTGLCDEKNA-VLERMEFPEVHIHAYE 412
QY 458 PENKNDLEKFSKIGRFTEDPTPKKYPTTENKETYISNGELHLEIYQRLEREKGC 517
DB 413 PTKVADQEKVVALGLAEEDPSFRVWTEEGQTLIGMGEHLHEIYDRKREPKVA 472
QY 518 ITGKRPVARETITTAVPDPFTHKQSGAGQYKVIYGLVLEPDPEDYTKLEFSDETFSS 577
DB 473 EIGQGVARETIRSSVSKENKAYKANGSGRGQYGHVPILEKREPS--SGYFVAVHISG 530
QY 578 NIPKQFVAREKGFPLDACERKPLSGHKLSGRFLVLDGAHNVDSNEISFIRAGECALQ 637
DB 531 VIPKEIYPAVDKGIQGMONGVLAGYVVDVFKVTLIDGSHVDVDSSEMAFKLAGNAFKE 590
QY 638 ALANATLCTLEPIMAVEVYAPNEFOQVYAGINRRGVITGQDGVEDYTYLADVPPLNM 697
DB 591 ASRAANPVLLPEPMKVEVPEBYMGDVIGDINRRGQJNSMDRLGLKTIKIVAFPLVEM 650
QY 698 FGYSITELRSTEGKEGYTMEYSRYQCLPSTOEDVINK 735
DB 651 FGYSITDLRSATQGRGTYSMEPDHYGEVPSNIAKEIYEV 688

```

```

RESULT 8
EFG_YERPE
ID EFG_YERPE STANDARD; PRT; 702 AA.
AC 0827B3;
DT 15-JUN-2002 (rel. 41, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR YPO0202.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI Taxid=632;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Bivort Orientalis;
MEDLINE=21470413; PubMed=11586360;
Parthill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
Baker S., Basham D., Bentley S.D., Brooks K., Cardeno-Tarraga A.M.,
Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
Fetwell T., Hamlin N., Holtroyd S., Uggels K., Kariyehov A.V.,
Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

```

```

RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AJ414141; CAC99063.1; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EF-G.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EF-G_C1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU D2; 1.
DR PRINTS; PR00315; ELONGATNFCF.
DR TIGRPFAM; TIGR00484; EF-G; 1.
DR TIGRPFAM; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 702 AA; 77537 MW; 8977FC9516165D2A CRC64;

```

```

Query Match 37.5%; Score 1462; DB 1; Length 702;
Best Local Similarity 44.4%; Pred. No. 5,7e-89;
Matches 317; Conservative 132; Mismatches 235; Indels 30; Gaps 15;

```

```

QY 42 PNEKIRNIGISAHIDSGKTTLTERVLYYTGRIAKNEHVKGDVGAVMDSMELEKRGIT 101
DB 6 PLNRIRNIGISAHIDAGKTTTERRILFYGVSHKIGEV--HDG-AATMDMMEQEKERGIT 62
QY 102 IQSAATFTFMKDVNINIIDTPGHVDFTTEVERARLVLDGAVLVLCVGVGVOQOTMVRNQ 154
DB 63 ITSAAITTCFWSGMKQEPFHHVNIIDTPGHVDFTTEVERARLVLDGAVLVLCVGVGVOQ 122
QY 155 TMTVROKRYNVPLFTINKLDRMGSNPAPALQOMRSKLNHTAFMOJPMGLEGNFKGI 214
DB 123 SETVROKRYNVPLFTINKLDRMGSNPAPALQOMRSKLNHTAFMOJPMGLEGNFKGI 182
QY 215 VDLIERAP-1YDGGFSQIVRYGEIPAEIRAAATDHRQELIECVANSDEQLGEMLEK 273
DB 183 IDLVKRAIKNNNEADQGTFTPEYBEIPAMAELEAEWQNLVESAARASDELMDKXLTPE 242
QY 274 PISIDLKLAIRATIKRSTPTPLGSLAKNGVOPLDVALEYLPNPEVQNYALINKKD 333
DB 243 LTBEEIKYKALRQVLSKSIILVTCGSAFKNGVQMDVALEYLPAPDVE--SINGILD 300
QY 334 DSKERTKILMNSSRH--NSHPFVGLAPFLEVGRF-GQLTYVRSYQGEIKKGDITYNTRR 390
DB 301 DSKDTPAV-----RHSDKEPFSALAPKIAIDPFVGNLTFRVYSGIYNSGDTVANSYS 355
QY 391 KKVRLQRLARWADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHVPD 449
DB 356 QRERLGRIVQNHANRE-EIKEVAGDIAAIGLKDVTGTGLCD-PNNPIILERMERPE 413
QY 450 PVISIAKMPKNDLEKFSKIGRFTEDPTPKKYPTTENKETYISNGELHLEIYQRLERE 509
DB 414 PVISIAVAPKTKVADQEKVVALGLAEEDPSFRVWTEEGQTLIGMGEHLHEIYDRKRE 473

```

QY	510	EEBVEGCPCTLTGPKVAFRETTIAPV-PFQTHKKSGAGGVGVGLVLEPDPEDYTKL	568
DB	474	RREPVANVAVGRPOVAVRETIETIVDVGKAKAGSGGQIGHVIMDSPLPGG-VGI	532
QY	569	EFSDETFGNSIKQVPAVEKGFLODACEKGPLSGHKLGLRFLVLDGAAHMYDSNEISFI	628
DB	533	EFVNEVVGSGISPEKEFFIPAVDKGIQEOCLKSGPLAGYFVVVVKVRLAHGSHYDVDSLEAF	592
QY	629	RAGEGLAKQALNATILCIIEPIMAVVAVNAFNEHQGVINGNRHGVITGQDSEVEYFTL	688
DB	593	LAGSLAFKGEFGKAKVLEPIPMKVEVEEPEDMGVMGDLNRRRIISMETATGKTV	652
QY	689	YADVLDNDFGXYSTELRSCTEGKGEYTWESRYQPCLPSTOEDVINKYLEATQ	742
DB	653	RKVPLSEMFVGTATDRSQTQGRASYSMEFLEYAAE-PS---NVAAVIEARCK	702
RESULT 9			
EFG_CAMUJE			
ID	EFG_CAMUJE	STANDARD;	PRT; 691 AA.
AC	Q9P16;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Elongation factor G (EF-G).		
GN	FUSA OR CJO493.		
OS	Campylobacter jejuni.		
OC	Campylobacter.		
CC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;		
CC	Campylobacter.		
CC	NCBI_TaxID=197;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NTC 11168;		
RX	MEDLINE=20150912; PubMed=1068204;		
RA	Parshall J., Wren B.W., Mungall K., Kealey J.M., Churcher C.,		
RA	Barham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,		
RA	Jegals K., Katlayeh A.V., Moulie S., Pallen M.J., Penn C.W.,		
RA	Quail M.A., Rajandream M.A., Rutherford K.M., van Vleet A.H.M.,		
RA	Whitehead S., Barrett B.G.;		
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni		
RT	reveals hypervariable sequences.";		
RL	Nature 403:665-668 (2000).		
CC	- FUNCTION: This protein promotes the GTP-dependent translocation of		
CC	the nascent protein chain from the A-site to the P-site of the		
CC	ribosome.		
CC	- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.		
CC	EF-G/EF-2 SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboratio		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on it		
CC	use by non-profit institutions as long as its content is in no wa		
CC	modified and this statement is not removed. Usage by and for commercia		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL, AL139075; CAB75131.1; --		
DR	HSSP, P13551; IELO.		
DR	InterPro: IPR004540; EF-G.		
DR	InterPro: IPR000640; EFG-G.		
DR	InterPro: IPR004161; EFTU D2.		
DR	InterPro: IPR000795; EF GTPbind.		
DR	InterPro: IPR005225; Small_GTP.		
DR	Pfam: PF00679; EFG_G, 1.		
DR	Pfam: PF00009; GTP_EFTU, 1.		
DR	Pfam: PF03144; GTP_EFTU D2, 1.		
DR	PRINTS; PR00315; ELONGATNCTC.		
DR	TIGRFAMs; TIGR00484; EF-G; 1.		
DR	TIGRFAMs; TIGR00231; small_GTP, 1.		
DR	PROSITE; PS00301; EFACOR_GTP, 1.		
DR	Elongation factor; Protein biosynthesis; GTP-binding;		
KW	Complete proteome.		

FT	NP BIND	17	24	GTP (BY SIMILARITY)	
FT	NP BIND	81	85	GTP (BY SIMILARITY)	
FT	NP BIND	135	138	GTP (BY SIMILARITY)	
SO	SEQUENCE	691 AA;	76719 MW;	16D3BF62B749CC60 CRC64;	
	Query Match	37.5%;	Score 1460.5;	DB 1;	Length 691;
	Best Local Similarity	44.3%;	Pred. No. 7e-89;		
	Matches	309;	Conservative 140;	Mismatches 230;	Indels 19; Gaps 11
QY		42	PNEKIRNIGISAHNIDSGKTTILTERLYYTGRIAKNHEVKGKDGVANWDSMLEROGIT	101	
DB		6	PLKVRNRNGIAAHIDAGKTTISERLIFPTGSHKIGEV--HGG-AATADMMQEKERGIT	62	
QY		102	IQSATFTPMADVNINIDTPEGHVFTTEVERALRLVGLAVLVLCAGVGVCQCTMTVNRQ	161	
DB		63	ITSATTCFMDHQNLDITPGHVDFTIEVERSMVLDGAAVAFSGVGVCQGSERVTNRQ	122	
QY		162	MKRYNVPLRTFINKLDRMGSNPARALQOMRSKLNNTAFMOJPMGLGNFKGIVDLIEER	221	
DB		123	ANKGVPRFIVEVNNKMDRIGANFYVNVEDDIRNLKANPRLQPIGAEDNFKGVIDLVYTK	182	
QY		222	AIYDGDPSQIVRY--GEIPAEIRAAADHQBOLIECVANSBOLGEMPLEEKISISRL	279	
DB		183	ALWVEDD-TKTDYDYEKEIPAELEKAEERTKMLEAVSETSDMEKYLGEESBLEEI	241	
QY		280	KLAIIRATLKRSPFVPLFGSALKKKGVORLLDAVLEYLPNPSVONYALINKKDSKEKT	339	
DB		242	KTGIRKAGCLSLSIYVMLCGTAFKKNQGVORLLDAVAVYLPAPDEVANI-----KGEYEDGT	296	
QY		340	KILNNSHNSHNPFGGLAFPLEVGRF-GOLTYVRSYQSELIKGGDTIYNTRTKRYRLQRL	398	
DB		297	EVSVKST--DGEFPAGLAFKIMTDPFVQLTFVRYRRCLESGSYAIVNSTKKKERIGRL	354	
QY		399	ARMHADMMEAETEEVYADICALFGI--CCASGDTFTDANSGLSMESIHPDPVISIAKK	457	
DB		355	LKMSNKRK-RIVKVLVAGEIGAVVGLKOTLIGDTLASKDK-VILERNDFPRVISIAVE	412	
QY		458	PSNRKDLKESFKIGRFTREDPTFKVYDTEKNETKTVISGMGLHLEIYAQRLEREYGCPC	517	
DB		413	PKTADQCKMSIAINKLQAQBDPSPFVSDSESGQTIISGMGLHLEIIVDRMLREKVA	472	
QY		518	ITGPKVAFRETITAPVPDPFTHKQSGAGQGVAVIGLBEJDPEDDYTKLEFSDPTFS	577	
DB		473	EVGQPOVAVRETIRKTVQEKRYAKQSGRGQYGVHFLRLEBFG--SGYFVANDIKGG	530	
QY		578	NIPKQFVAVKGFGLDACEKQPLSGHLSGLRFLYLDGAAHHMVDNSNISFIRAGECALQ	637	
DB		531	VIPREYIPAVDKGVQDALQNGVLAVGYPVEDYKVTYVDSYHEVDSSEMAFKLAASGFKE	590	
QY		638	ALANATLCLEPPIAVVEVAENEPQGVITAGINRHRGIVITQDGVGVEDYTYLADVPDLM	697	
DB		591	GARRAGAVILEPMMKVEVEPEDYEDVDVIGDINKRRGVONSDERGKNKIIITAFCPLEAM	650	
QY		698	FGYSTELRSCTEGKEVTEMSRYQCPSPSTQEDVINK	735	
DB		651	FGYSTDLRSQTQGRATYSMERDHDVKNNADELTKK	688	
RESULT 10					
EFQ_SPIPL	STANDARD;	PRT;	697 AA.		
AC	P13550;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Elongation factor G (EF-G).				
GN	FUSB OR FUS.				
OS	Spirulina platensis.				
OC	Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.				
OX	NCBI_Taxid=1156;				
RM	(1)				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89364697; PubMed=2505055;				

RA Butarelli F.R., Calogero R.A., Tiboni O., Gualerzi C.O., Pon C.L.;  
 RT "Characterization of the str operon genes from *Spirulina platensis*  
 and their evolutionary relationship to those of other prokaryotes";  
 RL Mol. Gen. Genet. 217:97-104(1989).  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION  
 OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 RIBOSOME.  
 CC SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X15646; CAA3672.1; -  
 CC DR PIR: S04390; S04390.  
 CC DR HSRP: P13551; 2EFG.  
 CC DR InterPro: IPR000450; EF-G.  
 CC DR InterPro: IPR000460; EFG\_C.  
 CC DR InterPro: IPR004161; EFTU\_D2.  
 CC DR InterPro: IPR000795; EF\_GTPbind.  
 CC DR InterPro: IPR005225; Small\_GTP.  
 CC DR Pfam: PF00069; GTP\_EFTU; 1.  
 CC DR Pfam: PF00679; EFG\_C; 1.  
 CC DR Pfam: PF0144; GTP\_EFTU\_D2; 1.  
 CC DR TIGRfam: TIGR00231; small\_GTP; 1.  
 CC DR TIGRfam: TIGR00484; EF-G; 1.  
 CC DR PROSITE: PS00301; EFATOR GTP; 1.  
 CC KM Elongation factor: Protein biosynthesis; GTP-binding.  
 FT NP\_BIND 17 24 GTP (BY SIMILARITY).  
 FT NP\_BIND 81 85 GTP (BY SIMILARITY).  
 FT NP\_BIND 135 138 GTP (BY SIMILARITY).  
 SQ SEQUENCE 697 AA; 76775 MW; 28C6DA1837DEB2C CRC64;

Query Match 37.3%; Score 1454; DB 1; Length 697;  
 Best Local Similarity 44.0%; Pred. No. 1.9e-88;  
 Matches 309; Conservative 18; Mismatches 234; Indels 22; Gaps 11;

QY 41 IPNKKIRNIGISAHIDSGKTTLLTERLYTGRIAKHEVKGQGVAVMSLELROSCI 100  
 DB 5 ILERVRNIGIAAHIDAGKTTTTERLTFYSGVHNKGEV--HSGT-AVIDMMAQERERKI 61  
 QY 101 TIQSAATFTMKQVNIINIIDTGHVDFTIEVERALRVNLGAVLVLCAGVCCQMTVNR 160  
 DB 62 TITAAAIISTSWLDHRINIIDTGHVDFTIEVERSMKRLVGLIAVFCVGVQPOSSTVWR 121  
 QY 161 QMKRVNVPFLTINKLDRKMSNPARALQOMRSGKLNNTNFMQIPMGLENFNGIYDLER 220  
 DB 122 QMKRVNVPFLTINKLDRKMSNPARALQOMRSGKLNNTNFMQIPMGLENFNGIYDLER 181  
 QY 221 RAIVFDGDFSQIVRYG-EIPAEIRAAATDRHOLEIECVANSDEQLEMLEE----KIPS 275  
 DB 182 KTYLVNTDGLDIOYSDIEPEVQDVAVYREKLEAVATEBALMEKLEQLEGEALAT 241  
 QY 276 ISDLKLAIRATLKRSFTVPVGLSALKNGVOPPLDVALEYLPNPEVONVAILNKDSS 335  
 DB 242 EEEIRHSLKQGTIKGLIVPVGICSSPKNGVORLLDAVDYAPAPTEVP--PIKGLVLPDQ 299  
 QY 336 KEKTIKLNSSRHNSHPFVGLAFLPELVGRFGLTYVRSVQGLKKGGDTIYNRTKRYNL 395  
 DB 300 EBGVRYADDA-----PLSALAFKVAADPYGRITFRVSVGLQKSYINATKNGKEKI 354  
 QY 396 QRLARHADMWASTEEVYAGDICALFGI--DCASGDTFTDKANSGLMSIHVPPIVSI 454  
 DB 355 SRLIVLKSD-ERIEVEBELAAGDGLAIGKDLITGTTIDENAS--IILESLVIREVISV 412  
 QY 455 AKPKSRNDLEKPSKIGFTBEDPFPKYFTTENKEIYISGMGELHLEIYQRLEREYX 514

DB 413 AVEPKTKDMKSLKALSLSEEDPTFRVSIIDSETNQTVIAGMGLHLEILVDRMLREFK 472  
 QY 515 CPCIRGKRVAFRETRITAPVPDFDTHKKQSGAGQYGVIGVLEPLDPEDYTKLEPSDET 574  
 DB 473 VEANIGAOVAVRETRISITREKGFIRQSGKQGYGVNVELEBGEFG--SGFFPVSKI 530  
 QY 575 FGSNIKQFVAVKEXFLDACERKGLSGRLSGRLVLDGAHHVMSNISPIRAGECA 634  
 DB 531 VGSVPEKVIYNDPAGMKACESGVIAGYPLIDVATLVDSYHEVDSSEMAFKIAGSMA 590  
 QY 635 LKQALANNTCLIEPIYAVNEVVAPEFGQVIAGINRHGVITQ--DGVEDYFLYADV 692  
 DB 591 IKNGYTKASPVLEPMKVEVEVPEDFIGNVIGDLSNRQIEGQETDQSGSIARVAKV 650  
 QY 693 PLNDVFGYSTELRSTCEKGEYTWESRYOCLPSTORADVINK 735  
 DB 651 PLATMFGYATDIRSKTGQRGVPSMESHVEEVPKSVAEITIAK 693

RESULT 11  
 EFG\_HAEIN  
 ID EFG\_HAEIN STANDARD; PRT; 699 AA.  
 AC P43525;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Elongation factor G (EF-G).  
 GN FUS1 OR FUS OR HT0579.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / K120 / ATCC 51907;  
 RX MEDLINE=9530630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merriek J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shriver C.A., Spriggs T., Hedblom R., Cotton M.D.,  
 RA Weisman J.F., Phillips R., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae*  
 Rd";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION  
 OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 RIBOSOME.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR Pfam; PF003144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATNFCF.  
 DR TIGRFAMs; TIGR00231; small GTP; 1.  
 DR TIGRFAMs; TIGR00484; EF-G\_1.  
 DR PROSITE; PS00301; EFATOR\_GTP; 1.  
 DR Elongation factor; Protein biosynthesis; GTP-binding;  
 KW Complete proteome.  
 FT INIT MET 0 BY SIMILARITY.  
 FT NP\_BIND 16 23 GTP (BY SIMILARITY).  
 FT NP\_BIND 87 91 GTP (BY SIMILARITY).  
 FT NP\_BIND 141 144 GTP (BY SIMILARITY).  
 SQ SEQUENCE 699 AA; 77132 MW; FBBAD63C0F62801 CRC64;  
 Query Match 37.3%; Score 1453; DB 1; Length 699;  
 Best Local Similarity 45.1%; Pred. No. 1,9e-88;  
 Matches 318; Conservative 124; Mismatches 235; Indels 28; Gaps 13;  
 QY 42 PNEKRNIGISAHIDSGKTTLTERTVLTGRARAKHEVKGKDGAVGAVMDSMELERQGIT 101  
 DB 5 PIERVNNIGISAHIDAGKTTTERTLTFTGVSHKIGEV--HDG-AATMDMEQEQERGRT 61  
 QY 102 IGSATFTMMKGVN-----NINIDTGGHVDFTIEVERALRVLDGAVLVLCAGVGVOQ 154  
 DB 62 ITSATFTAFSGSQOPQGRINIVIDTGGHVDFTIEVERSMRVLDGAVLVLCAGVGVOQ 121  
 QY 155 TWTVNRQMKRYNVPFLTINKLDRMGSNPARALQOMRSKLNHTAFMOI PMGLEGNFKGI 214  
 DB 122 SEVWQANKYEVPRALAFVNMKDRGTGANFLRVGEQLKTRGANAIIQLQPVGAENFTV 181  
 QY 215 VDLIERALYFGDGDSQIVRYGEI PAELRAAATDHRQELIECVANSDEQLGEMFLEEK 273  
 DB 182 VDLIKKAIWNADQGMFTYEVPANMQADCEBRQNLVEAAASAEELMEKYLGGED 241  
 QY 274 PSISDGLAIRATLRKSPFVFLGSAIKKGVQPLDALTXYLPMPSEVQNAIILKKD 333  
 DB 242 LTREELKSAIRQVLANEIIIVTCSAFKKGVQAMLDAAVEYLPAPTDLP--AIKGINP 299  
 QY 334 DSKEKTKILNNSRNHS--HPVGLAFPLEVGRF--GQLTVRSYOGELKKGDTTYNTRR 390  
 DB 300 DETE-----GERHNSDESPFSLAFKLTATDPFQNLFFFRVYSGVINSGLDVLNSVQ 352  
 QY 391 KATRLQRLARMAADMEASTEEVYADICALFPI--DCASGDTTDXANSGLSMESIHVPD 449  
 DB 353 KRERFRIVOMHANKKE--EIKERAGDIAAIIKQVLTGDTLC--AIDAPILIERMEFPE 410  
 QY 450 PVSIAKPKENKNDLEKFSKIGRFRBDPTFKVYFOTENKETVISGMELHLEIYAOL 509  
 DB 411 PVSIVAVEPTKADQEKMGIALGRLAQEDSFVHTDESGEITISGMELHLDITVDRK 470  
 QY 510 EREYGCPTIGKPKVAFRITITAPV--PDPFTHKQSGAGQXGVIGVLEPLDPEDYTKL 568  
 DB 471 KRBFKEANIGKQVSRFETIRTRVNDVEGKAQSGRQYGHVVIDYPLDPEG--PGV 529  
 QY 569 EREDETFGSIPIQOPFPAVEKGLDACCKPLSGHKLGLRFLQDGAHMDVDSNISTFI 628  
 DB 530 EFNELKGGVIPGEYIPAVDKGIOEQLKSGPLAGYVVDIGVLAHSGSYNDVDSSELARK 589  
 QY 629 RAGEGALKOALNAATCILEPIMAVEVAPNEFOGOVIGINRHGVIITGODVEVEYFL 688  
 DB 590 LAAASLAFKAAFSQANFVLEPIIMKEVEYTPPEYVGVIDGLSSRRRAMVNGQEANBEVVKI 649  
 QY 689 YADVPLNDMGYSTELRSCTEGKGEYTMEXSYRQPCLPSTOEDVI 733  
 DB 650 YAEVPLSEMGVATDLSQTOGRASYSMEPLKTAEAAPTVAALVI 694

DE Elongation factor G 2 (EF-G 2).  
 GN FUSB OR FUSA2 OR RSP0804 OR RS06124.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC plasmid megaplasmid  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_Taxid=305;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoudat M., Genin S., Artiguenave F., Gouzy J., Mangerot S.,  
 RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Siegler P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of  
 the nascent protein chain from the A-site to the P-site of the  
 ribosome.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL646081; CAD17955.1; -  
 DR InterPro; IPR004540; EF-G.  
 DR InterPro; IPR000640; EF-G.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR005225; Small\_GTP.  
 DR Pfam; PF00679; ERG\_C\_1.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATNFCF.  
 DR TIGRFAMs; TIGR00484; EF-G\_1.  
 DR TIGRFAMs; TIGR00231; small GTP; 1.  
 DR PROSITE; PS00301; EFATOR\_GTP; 1.  
 KW Elongation factor; Protein biosynthesis; GTP-binding; Plasmid;  
 KW Complete proteome.  
 FT NP\_BIND 17 24 GTP (BY SIMILARITY).  
 FT NP\_BIND 88 92 GTP (BY SIMILARITY).  
 FT NP\_BIND 142 145 GTP (BY SIMILARITY).  
 SQ SEQUENCE 700 AA; 77619 MW; 53CC4CA877EE919 CRC64;  
 Query Match 37.3%; Score 1453; DB 1; Length 700;  
 Best Local Similarity 45.2%; Pred. No. 2,2e-88;  
 Matches 320; Conservative 126; Mismatches 230; Indels 32; Gaps 14;  
 QY 42 PNEKRNIGISAHIDSGKTTLTERTVLTGRARAKHEVKGKDGAVGAVMDSMELERQGIT 101  
 DB 6 PIERVNNIGISAHIDAGKTTTERTLTFTGVNHKIGEV--HDG-AATMDMEQEQERGRT 62  
 QY 102 IGSATFTMMKGVN-----NINIDTGGHVDFTIEVERALRVLDGAVLVLCAGVGVOQ 154  
 DB 63 ITSATHTCMKMGNGYVEHRIINIIDTGGHVDFTIEVERSMRVLDGACMVDSVSGVQPO 122  
 QY 155 TWTVNRQMKRYNVPFLTINKLDRMGSNPARALQOMRSKLNHTAFMOI PMGLEGNFKGI 214  
 DB 123 SEVWQANKYEVPRALAFVNMKDRGTGANFLRVGEQLKTRGANAIIQLQPVGAEDHFRGV 182  
 QY 215 VDLIERALYFGDGDSQIVRY--YGEI PAELRAAATDHRQELIECVANSDEQLGEMFLEEK 272  
 DB 183 VDLVKKAIWVD--DASQGVKEFYIDIPEELRATQAEWHDKNVEAAAEADALLEKYLGE 241





AC P57593;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Elongation factor G (EF-G).  
 GN F05A OR B0527.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 OS symbiotic bacterium).  
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
 OX NCBI\_TaxId=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RA MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT Genome sequence of the endocellular bacterial symbiont of aphids  
 RT Buchnera sp. AS-".  
 RL Nature 407:81-86(2000).  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION  
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 CC RIBOSOME.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF001119; BAB13220.1; -.  
 CC HSSP: P13551; 1ELO.  
 DR InterPro: IPR004540; EF-G.  
 DR InterPro: IPR000640; EF-G.  
 DR InterPro: IPR004161; EFTU D2.  
 DR InterPro: IPR000795; EF GTPbind.  
 DR InterPro: IPR005225; Small GTP.  
 DR Pfam: PF00009; GTP\_EFTU; 1.  
 DR Pfam: PF00679; EF-G\_C; 1.  
 DR Pfam: PF03144; GTP\_EFTU\_D2; 1.  
 DR Pfam: PF00315; ELONGATINCT.  
 DR PRINTS: PR00231; small GTP; 1.  
 DR TIGRPFAM: TIGR00484; EF-G\_1.  
 DR TIGRPFAM: TIGR00301; EFATOR\_GTP; 1.  
 DR Elongation factor; Protein biosynthesis; GTP-binding;  
 KM Complete proteome.  
 FT NP\_BIND 17 24 GTP (BY SIMILARITY).  
 FT NP\_BIND 88 92 GTP (BY SIMILARITY).  
 FT NP\_BIND 142 145 GTP (BY SIMILARITY).  
 SO SEQUENCE 702 AA; 78182 MW; F583FEP269ABICF9 CRC64;  
 Query Match 36.9%; Score 1438; DB 1; Length 702;  
 Best Local Similarity 44.5%; Pred. No. 2, 2e-87;  
 Matches 317; Conservative 126; Mismatches 241; Indels 28; Gaps 14;

QY 270 EKKIPISDLKLAIRATIKRSFTVPFPGSALKKNGKOPLDVALEYLPNPEVQNT-AI 328  
 DB 229 NODRISEBEIISALKRALNNEIVILITGSAFKKNGVALDLAIEFLPANDIDINGI 298  
 QY 329 LKKKDSKERTKILNNSRRHNSHPVGLAFLPLEVGRF-GQLTYVRSYGELKGDITYNT 387  
 DB 239 LN----DVEQTPAIRNSD--DKAPPSALAFKIASDPFGNLTFFRVSGVKSQDTPNS 352  
 QY 388 RTRKKVRLQRLARMADMBASTEEVAGDICALFGI--DCASGDFFTKANSGLSMESI 446  
 DB 353 AKSORERGRIVQMPANKRE-EIKSVYAGDIAAIGLADVTGDTLCD-LNDPIILERME 410  
 QY 447 VDDPVISIAIKPSNRNDLEKESKIGREFREDPTFKYVFDENKETVISGMGELHLEIYA 506  
 DB 411 PEPVVISISVEKTKVDEKNGKALGRALKEDPSFRVNTDESNTIISGMGELHLEI 470  
 QY 507 QRLEREYCPCTIKPKYAFRETTAPV-PEDFTHKKSGAGQYKXIVLEPLDP--E 563  
 DB 471 DMKKEFVSADVAGKQVAYRETTINKVEDIEGKIKOSGGQYGHVIELFPLQPGE 530  
 QY 564 DYTKEFDEFTGSIPIKQFVPAVEKGLDACEKGLSGHKLSGRLFVLDGAAHMDVN 623  
 DB 531 GYL--FVNDIKGVIPSEYISAIKGIQEDKCGPLAGVPVDIGVRLYRSYHVDVS 587  
 QY 624 EISFIRAGEGALKQALANATLCILEPINAVEVPAFNEFGOVIAGINRRHGVIITQDVE 683  
 DB 588 ELAFLTAASAAPFKQFKQAKPLLEPIKMEVEETPDYDVGIVDGNRRRGIEGMKOLS 647  
 QY 684 DYFTLYADVPPLDMMGISTELBSCTEGKGTMEISRQPLCSIQEIVYINK 735  
 DB 648 ISKINACVPLESEMPGATDLRSOTQGRASYSMEFLKYIEAPFNISKDIIR 699  
 RESULT 15  
 EFG\_SALTY STANDARD; PRT; 703 AA.  
 ID EFG\_SALTY  
 AC P26229;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Elongation factor G (EF-G).  
 GN FUSA OR STM3446 OR STY4352.  
 OS Salmonella typhimurium, and  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxId=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. typhimurium; STRAIN=LT2;  
 RX MEDLINE=93013025; PubMed=1398129;  
 RA Johnson H.U., Hughes D.;  
 RT typhimurium and Escherichia coli.";  
 RL Gene 120:93-98(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. typhimurium; STRAIN=LT2 / SGC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvoney E.,  
 RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,



Query Match	36.8%	Score 1434	DB 1	Length 703
Best Local Similarity	44.2%	Pred. No. 4e-87		
Matches	316	Conservative	126	Mismatches 243; Indels 30; Gaps 14
Qy	42	PNEKIRNIGISAHIDSGKTTLTERTVLYTGRIAKHEVKKDGVAMDSMELERQGIT	101	
Db	5	PIARRNRNIGISAHIDAGKTTTERTLTFTYGVHKHIGEV--HDG--AATWDMMEQBERGIT	61	
Qy	102	IOSAATFTMMMDV-----NINIDTGFHVDFTIEVERALRVLDGVALVCAVGGYCO	154	
Db	62	ITSAATTAFWGSGMAQYEPHRNINIDTGFHVDFTIEVEREKRVLDGVAWVYCAVGGVPQ	121	
Qy	155	TATVNRQMKRNVNPLFTPIINKLDRNGSNPARALOQMRSLNHNATFWQIPMGLEGNFYGI	214	
Db	122	SETVNRQANKKVPRIAFVNRKQDRNGANFLKVVQGIKTRIGANVPVLTAIGAEFGFVG	181	
Qy	215	VDLIEERAI-YEDGPGSQIVRYGSEPAELRAAATHROEILTECVANSDEOIGEMLEEKI	273	
Db	182	VDLVKMKILNNNDACQGYTFEYEDTIPADMODLANEMHQNLIESAAEABEELMEKTLGSEE	241	
Qy	274	PSISDLKAIIRRAITLRSGFTPVFLGSAALKKNGVQPLDVALEYLENPSEVONYATILNKD	333	

Search completed: July 14, 2003, 18:16:10  
Job time : 10.8217 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 / Search time 16.6366 Seconds  
(without alignment)  
4345.419 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898

Sequence: 1 MRLGMAAVALGRGRAPAS.....INKYLHATGQLPVKKKAKN 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3243	83.2	752	1 S40780	translacion elonga
2	2119.5	54.4	750	2 T21534	hypothetical proce
3	1987	51.0	761	1 S61642	translacion elonga
4	1972	50.6	754	2 D96510	probable mitochond
5	1970	50.5	754	2 P84885	hypothetical prote
6	1903.5	48.8	770	2 T50308	probable translati
7	1765	45.3	695	2 D71283	translacion elonga
8	1762	45.2	693	2 C70167	translacion elonga
9	1493	38.3	694	1 S04429	translacion elonga
10	1489	38.2	692	2 G71847	translacion elonga
11	1487	38.1	692	2 C64669	translacion elonga
12	1472	37.8	692	2 AC2348	translacion elonga
13	1462	37.5	702	2 AD0025	translacion elonga
14	1460.5	37.5	691	2 H81394	translacion elonga
15	1454	37.3	697	2 S04390	translacion elonga
16	1454	37.3	700	2 P64078	translacion elonga
17	1438	36.9	702	2 D84991	translacion factor
18	1434	36.8	704	2 JCI424	translacion elonga
19	1434	36.8	704	2 AC1005	translacion elonga
20	1431.5	36.7	691	1 EPTMG	translacion factor
21	1430	36.7	701	2 E70827	translacion elonga
22	1427	36.6	704	1 EPECG	probable fusa prot
23	1427	36.6	704	2 G91152	translacion elonga
24	1427	36.6	704	2 C85998	translacion elonga
25	1426.5	36.6	691	2 S75863	translacion elonga
26	1422	36.5	701	2 H87143	translacion factor
27	1416.5	36.3	701	2 B82006	translacion elonga
28	1416	36.3	701	2 S31150	translacion elonga
29	1414.5	36.3	694	2 AD3346	protein translatio

30	1413.5	36.3	701	2 C81234	translacion elonga
31	1408.5	36.1	705	2 P82534	translacion elonga
32	1407.5	36.1	695	2 H72243	translacion elonga
33	1406	36.1	702	2 G83386	translacion elonga
34	1401	35.9	687	2 A87286	translacion factor
35	1399.5	35.9	688	2 S73930	translacion elonga
36	1398.5	35.9	699	2 A70300	translacion elonga
37	1398	35.9	699	2 E97594	translacion factor
38	1398	35.9	699	2 AC2816	translacion elonga
39	1398	35.9	699	2 E97594	translacion elonga
40	1395	35.8	695	2 S76751	translacion elonga
41	1395	35.8	698	2 C82332	translacion elonga
42	1392.5	35.7	692	2 B69628	translacion elonga
43	1388	35.6	695	2 A81762	translacion elonga
44	1387	35.6	695	2 A81406	translacion elonga
45	1387	35.6	699	2 B71723	translacion elonga

## ALIGNMENTS

## RESULT 1

S40780

translacion elongation factor EF-G, mitochondrial - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001

C:Accession: S40780

R:Barker, C.; Makris, A.; Patriotic, C.; Bear, S.E.; Tichlis, P.N.

Nucleic Acids Res. 21, 2641-2647, 1993

A>Title: Identification of the gene encoding the mitochondrial elongation factor G in ma

A:Reference number: S40780; MUID:93324327; PMID:8332461

A:Accession: S40780

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-752 <BAR>

A:Cross-references: EMBL:L14684; NID:G310101; PID:AAA1107.1; PID:G310102

C:Genetics:

A:Map position: 2

C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

C:Keywords: GTP binding; mitochondrion; nucleotide binding; P-loop; protein biosynthesis

F:48-178/Domain: translation elongation factor Tu homology <Etu>

F:54-61/Region: nucleotide-binding motif A (P-loop)

F:175-178/Region: GTP-binding NKXD motif

F:300-302/Region: GTP-binding SKX/L motif

F:60,61,102,175,176,178,300/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #sc

Query Match	Score	DB 1:	Length	DB 2:	Matches	Conservative	Mismatches	Indels	Gaps	#sc
1	3243	83.2%	752	1	636	37	75	8	5	
1	MRLGMAAVALGRG---	RAPASLGWOKVNMWACRWSGCV	PNEKIRNIGSAHDS	57						
1	MRL--RTTAGLGRPLPRVPAII	GQKMMWTKYRCSSGS	IPNEKIRNIGSAHDS	58						
58	GKTLTERLVLYTGRIAMHEVK	KDGVAMDSMELEKRGITIGSAAT	ATPMKDVNIN	117						
59	GKTLTERLVLYTGRIAMHEVK	KDGVAMDSMELEKRGITIGSAAT	ATPMKDVNIN	118						
118	IIDPGRHVDFTIEVERALRV	LDGAVLVICAVGVQCOTMTVNR	KRYNVPPLFINKD	177						
119	IIDPGRHVDFTIEVERALRV	LDGAVLVICAVGVQCOTMTVNR	KRYNVPPLFINKD	178						
178	RMGSPAPALQMSKLNHNAF	POIIPMGEGNFKGIVDLIEE	AIYDGFSGQIVRGE	237						
179	RMGSPAPALQMSKLNHNAF	POIIPMGEGNFKGIVDLIEE	AIYDGFSGQIVRGE	238						
238	IPAEIRAAATDHRQELIECV	ANSDEQGEMLFEKISISDLK	AIRRAATKRSFTPVFL	297						
239	IPADIRAAAHREELIECV	ANSDEQGEMLFEKISISDLK	AIRRAATKRSFTPVFL	298						
298	GSALKNKGVPLDLAVIEYL	PNPSEVQNYAIIINKKDS	KEKTKILVNSSRNSHPVGLA	357						
299	GSALKNKGVPLDLAVIEYL	PNPSEVQNYAIIINKKDS	KEKTKILVNSSRNSHPVGLA	357						

QY 358 FPLEVGRFQQLTVRSYQOGLKKKDDTYNTRTKRYLRLARMAHMAEASTEEVYAGD 417  
DB 358 FPLEVGRFQQLTVRSYQOGLKKKDDTYNTRTKRYLRLARMAHMAEASTEEVYAGD 416  
QY 418 ICALFGIDCAGDFTDKANSGLSMESIHVPDPVISIAKPSNKDLEKFSKIGFTRE 477  
DB 417 ICALFGIDCAGDFTDKANSGLSMESIHVPDPVISIAKPSNKDLEKFSKIGFTRE 476  
QY 478 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGKPKVAARETTTAVPPD 537  
DB 477 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGKPKVAARETTTAVPPD 536  
QY 538 FTHKOSGGA-GQYKGYIVLEPLDEPDYTKLEFSDETSGSNIPKOFPAVEKGFLDACE 596  
DB 537 ITHKOSGGA-GQYKGYIVLEPLDEPDYTKLEFSDETSGSNIPKOFPAVEKGFLDACE 596  
QY 597 KGPLSGHKLGLRFLVLDGAHMHVDSNEISFIRAGEGALKQALANATLCILEPIMAVEV 656  
DB 597 KGPLSGHKLGLRFLVLDGAHMHVDSNEISFIRAGEGALKQALANATLCILEPIMAVEV 656  
QY 657 APNEFGQAVFAGINRRHGVITGDDGVEDFTLYADVPLDMFGYSTELASCTEGKEEYTM 716  
DB 657 APNEFGQAVFAGINRRHGVITGDDGVEDFTLYADVPLDMFGYSTELASCTEGKEEYTM 716  
QY 717 EYSRYOPLPSTQEDVYINKYLEATGQLPYKKGAKN 752  
DB 717 EYSRYOPLPSTQEDVYINKYLEATGQLPYKKGAKN 752

## RESULT 2

T21534  
hypothetical protein F29C12.4 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
C:Accession: T21534  
R:Steward, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19436  
A:Accession: T21534  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-750 <WIL>  
A:Cross-references: EMBL:Z81519; PDB:CA04216.1; GSPDB:GN00020; CESP:F29C12.4  
A:Experimental source: clone F29C12  
C:Genetics:  
A:Gene: CESP:F29C12.4  
A:Map position: 2  
A:Introns: 16/3; 48/1; 287/3; 438/3; 531/2  
A:Superfamily: translation elongation factor G; translation elongation factor Tu homolog  
F:45-175/Domain: translation elongation factor Tu homolog <ETU>

Query Match 54.4%; Score 2119.5; DB 2; Length 750;  
Best Local Similarity 54.9%; Pred. No. 2.5e-130;  
Matches 414; Conservative 125; Mismatches 198; Indels 17; Gaps 6;

QY 9 VALGRGRAPASLGWQKQVNMKACRMS-----GVTPNEKIRNIGISAHIDSGK 59  
DB 1 WSLGKSPSAVQ--SRQDANVLRFPASNEASVVPGRVTEIRIRNIGISAHIDSGK 57  
QY 60 TLLTERVLYYTGRIAKMHEVKGQGVAVWDSMELEROGITIOSAATPTMMKDVNINI 119  
DB 58 TLLTERVLYYAGRIDSMHEVRGDDVGAITMDPDLRQGITIOSAATVDMHGNINII 117  
QY 120 DTPGHVDFITTEVERALRVLDGAVLVICANGVCCQMTNRRQKRYNVPPLTFINKLDM 179  
DB 118 DTPGHVDFITTEVERALRVLDGAVLVICANGVCCQMTNRRQKRYNVPPLTFINKLDM 177  
QY 180 GSNPARALOQMSKLNHNTAFMOIPMGLEGNFKGYIDLIEBRIYDGDGFSQIVRYGEIP 239  
DB 178 GATPLKALDGLRKLKLNHNAALHLPLPGKSNFNGIYDLVBGHALYEGSGGLVRCDEIP 237  
QY 240 AELRAAATDHRQELIECVANSDEQLGEMPLEEKIPISIDLKLAIRPATKRSFTFVLGS 299

DB 228 KDLRVAEADRRQELIEHIANVDETLGEMFLNDQFNVQOIHIALIRRTVVKAFVPLVSGS 297  
QY 300 ALKXGVPLDVALEVIIPNPSVQNVAILINKKDSKEKTKILNMSRHNSHPGLAF 359  
DB 298 ALKXGVPLDVALEVIIPNPSVQNVAILINKKDSKEKTKILNMSRHNSHPGLAF 356  
QY 360 LEVGRFQQLTVRSYQOGLKKKDDTYNTRTKRYLRLARMAHMAEASTEEVYAGD 418  
DB 357 LEVGRFQQLTVRSYQOGLKKKDDTYNTRTKRYLRLARMAHMAEASTEEVYAGD 414  
QY 419 CALFGIDCAGDFTDKANSGLSMESIHVPDPVISIAKPSNKDLEKFSKIGFTRED 478  
DB 415 CALFGIDCAGDFTDKANSGLSMESIHVPDPVISIAKPSNKDLEKFSKIGFTRED 474  
QY 479 PFRKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGKPKVAARETTTAVPPD 538  
DB 475 PFRKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGKPKVAARETTTAVPPD 534  
QY 539 FTHKOSGGA-GQYKGYIVLEPLDEPDYTKLEFSDETSGSNIPKOFPAVEKGFLDACE 598  
DB 535 FTHKOSGGA-GQYKGYIVLEPLDEPDYTKLEFSDETSGSNIPKOFPAVEKGFLDACE 594  
QY 599 PLSGHLKSLRFLVLDGAHMHVDSNEISFIRAGEGALKQALANATLCILEPIMAVEV 658  
DB 595 PLSGHLKSLRFLVLDGAHMHVDSNEISFIRAGEGALKQALANATLCILEPIMAVEV 654  
QY 659 NEFGQAVFAGINRRHGVITGDDGVEDFTLYADVPLDMFGYSTELASCTEGKEEYTM 718  
DB 655 NEFGQAVFAGINRRHGVITGDDGVEDFTLYADVPLDMFGYSTELASCTEGKEEYTM 714  
QY 719 EYSRYOPLPSTQEDVYINKYLEATGQLPYKKGAKN 751  
DB 715 EYSRYOPLPSTQEDVYINKYLEATGQLPYKKGAKN 748

## RESULT 3

translation elongation factor EF-G, mitochondrial - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein L2195; protein YLR069c  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001  
C:Accession: S61642; S20179; S64897; S64901; S17025  
R:Ureastarzu, L.A.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: S61642  
A:Accession: S61642  
A:Molecule type: DNA  
A:Residues: 1-761 <URR>  
A:Cross-references: EMBL:X94607; NID:g1181264; PDB:CAA64315.1; PID:g1181289  
R:Vanburen, A.; Ackerman, S.H.; Tzagoloff, A.  
Eur. J. Biochem. 201, 643-652, 1991  
A:Title: Mitochondrial translational-initiation and elongation factors in *Saccharomyces*  
A:Reference number: S20177; MUID:92037620; PMID:1935960  
A:Accession: S20179  
A:Molecule type: DNA  
A:Residues: 1-65, '1', 67-232, 'V', 234-477, 'S', 479-628, 'P', 630-761 <VAM>  
A:Cross-references: EMBL:X58378; NID:g31917; PDB:CAA41267.1; PID:g31918  
R:Andre, B.; Ureastarzu, L.A.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64872  
A:Accession: S64897  
A:Molecule type: DNA  
A:Residues: 1-761 <AND>  
A:Cross-references: EMBL:Z73241; NID:g1360421; PDB:CAA97626.1; PID:g1360422; GSPDB:GN00  
R:Polh, T.M.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64899  
A:Accession: S64901  
A:Molecule type: DNA  
A:Residues: 1-761 <POH>  
A:Cross-references: EMBL:Z73241; NID:g1360421; PDB:CAA97626.1; PID:g1360422; GSPDB:GN00

A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:MEF1, MIPS:YLR069C  
A:Cross-references: SGD:S0004059, MIPS:YLR069C  
A:Map position: 12R  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog  
C:Keywords: GTP binding; mitochondrion; nucleotide binding; P-loop; protein biosynthesis  
F:71-205/Domain: translation elongation factor Tu homolog <ETU>  
F:77-84/Region: nucleotide-binding motif A (P-loop)  
F:202-205/Region: GTP-binding NKXD motif  
F:327-329/Region: GTP-binding SAK/L motif

Query Match	51.0%;	Score 1987;	DB 1;	Length 761;
Best Local Similarity	56.2%;	Pred. No. 1.1e-121;		
Matches 392;	Conservative 117;	Mismatches 172;	Indels 16;	Gaps 8

Qy 45 KIRINIGSAHIDSKETTLTERLVYTGIAHBEVKGQVGAVMSDEHELRQGITIOS 104  
Db 69 KLRNIGSAHIDSKETTLTERLVYTTKRIKAHHEVRGDNVGAQMSDMLEREKGIITIOS 128  
Qy 105 AATTTMM---KDVNINIIDTPGHVDTIEVERALRVLDGAVLVCAVGVOQOTMTVNR 160  
Db 129 AATYCSMDKECKNHFMLIDTPGHIDFTIEVERALRVLDGAVLVCAVSGVOQOTVTVDR 188  
Qy 161 QMKRYNVPFLFFIKCLDMSGNSPARALOQMSKJLNHNTAFQIIPMGEGNKGIVLDIEE 220  
Db 189 QMKRYNVPXYTFINKMDRMSGDPFRAIEQLSKSKITPAAYVQITIGSESSISLGSVVDLNR 248  
Qy 221 RAIYFDGDFSQIVAYGEIPAELEAAATDHRQELIECVANSDEQGFEMFLEBKIPSIDLK 280  
Db 249 VAIYKNGNGEIIERKGPVBNIKPLEBEKROLTIETLADVDENAMFLEBEKEPTQOIK 308  
Qy 281 LAIRATLTKSFTBEVFGSALJONGVOPLDANVLEVPNPEQONVYAILNKDSDSKERTK 340  
Db 309 DAIRRTIARFETVLKMSALANTGIDPVLDAIVDLYPNPSEVNTAL---DVSNNEAK 364  
Qy 341 ILNMSRHNHPFVGLAFPLEVGRFGQLTVRSYQGBELKKDDTYNTRTKRYVALORLAR 400  
Db 365 V--NLVPAVQOPVGLAFKLEBKKGQOLTYYRVYQGRLRKKNYLTNVTGKKVAVARLVR 422  
Qy 401 MHADMENSTEVEVYAGDICALFGIDCASGDTFTDKANSGLSMESIHVPDPYISIAMKPSN 460  
Db 423 MHSEME--DVDEVSGEICATFGIDCASGDTFT--GSVQYMSMAYPDAVAVSISTIPNS 480  
Qy 461 KNDLEKESKIGRFTREDPTFKVYFDTENKETVISINGMELHELEYAORLEREYSCPCITG 520  
Db 481 K-DASNSKALNRQKEDPTFRVAKFPDESKETIISGMGELHELEYVERMRREIVNDCVTG 539  
Qy 521 KPVAARETITAPVPDFPTHKKQSGAGQYKVGIVLEPLDPEDYTKLE--PSDETFSNI 579  
Db 540 KPQVSYRESITIPADPDFYTHKKQSGAGQYGRVIGTSLPVD--DITKGNIFETAIYGGRI 597  
Qy 560 PKQVVPVPEKGFLLACCEKGPLSGHKLSGLRVLDQGAHMHWSNEISPIRAGEBALQAL 639  
Db 598 PDKLLAACGCKFEERCEKGPLIGHRVLDVKPLINDGALHADVSNELSKTATMSAFRDAF 657  
Qy 640 ANATLTCLIEPIMAEVVAAPNEFOQVIAGIRRRRGVITGCQGVDDFTLYADVLANMFG 699  
Db 658 LRAQPVIMEPIMNVSVPNEFOQVIGLANKLQAVIQDTENGHDEFTLKACALSTMFG 717  
Qy 700 YSTELRSTBEKGEYTMYSRYOPLSTQGDVYNKY 736  
Db 718 FATSLRSTQCKGEFSLFESHYAPTAPHVQCKLSEF 754

RESULT 4  
D96510  
Probable mitochondrial elongation factor [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: D96510  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federepial, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A; Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani, Pizzotti M.; Rooney T.; Powlay D.; Sakano H.

A:Authors: Salzberg, S.L.; Schwartz, J.B.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, Rizzo, M.; Rooney, L.; Rowley, D.; Sakano, H.

A./HACHIOBI: Saitenberg, S.D.; Schmalzer, C.K.; Shihui, F.; Souciwack, A.M.; Sun, H.; Yalton,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: D96510

A; Status: preliminary

A;Molecule type: DNA

A;K residuals: 1 - /54 <5.1

C:Genetics: A;C1038-References: 9B:AE003173; MID:gl1321780; PIDN:AA034237.1; 3SPDB:GN0014

A;Map position: 1

C;Superfamily: trans1

**C; Keywords:** mitochondria

1. The first group of respondents (Group 1) consisted of 100 individuals who were randomly selected from a list of all employees of the company. This group was surveyed in the first quarter of 2018.

Query Match

Best Local Similarity 54.8%; Pred. No. 1.1e-120;  
 Matches 385; Conservative 118; Mismatches 187;  
 Indels 13; Gaps 5

```

Oy 35 WSSGVL PNEKIRINIGSAHIDSGKTLTEVLVYTRIAHMBGVKDGGAIVWDSMEL 94
Db 58 WKES----MDKRNIGSAHIDSGKTLTEVLVYTRIRIHEIHEVRBDGVGAKMDSMDL 113
Oy 95 ERORGITIOSAATFTMKVDVINIIDTPGHVDFTIEVERALVLDGVLVLCAVGVCQ 154
Db 114 EREKGITIOSAATFTMKVDYKVINIIDTPGHVDFTIEVERALVLDGAILVLCSVGVSQ 173
Oy 155 TMTNROMKRNVEFLPIFKLDMGSPARALQOMSKINHNTAPMOIPWGLEGNFGCI 214
Db 174 SITYDROMKRNVEFLPIFKLDMGADPMWVLVLOAKIRHNSAAVQVPLGLEBNFGCL 233
Oy 215 VDLIEBRALFYDGFDSQIVRYGEIPAEILRAATTHROBLIECVANSDEOLGEMLEEKIP 274
Db 234 IDLIHVAYAFPHSGSGENVVAGDIPADMEGLVAEKRELIETVSEVDVLAKEFLNDEPV 293
Oy 275 SISLKLAIIRATILKASFTPVFLGSALKNGVQPLDLAVLEYLNPSPSVONYAIALNKDD 334
Db 294 SASLEBRALRATIAQFTVPVFMGSAFNKNQVQPLDLGVVSFLPSPEVNNTAL---DQ 349
Oy 335 SKERTIKLIANSSRNSHSPVGLAEPLLEVGRFGQLTYYRSVQGLKKGDTIYNTRRKVR 394
Db 350 NNNEERTILGSGPBG--PLVALAKLEBGRFGQLTLYRYBVGVIKKGDFIINNATGRKIK 407
Oy 395 LQRLARNHADMEASTEEVYAGDICALFGIDCASGDTFTDKANSGLSNESITHVDPVISI 454
Db 408 VPLRYVRHMSNDME-DIOEHAAGQIVAVFGIECASGDTFTD-GSYKYTWTSMNVBEPWSL 465
Oy 455 AMKSNNDLEKPSFGIGRFTREDPTFKVYFDTEKNETVLSGMWELHLEYIAOGLBERPG 514
Db 466 AVQGVSDOSGGQFSKALNRFQKEDPTFRVGLDPEBSGQITISGMELHLDIYVERNRREYK 525
Oy 515 CPCITGPKVYAFRETTIAPVPFDPFTHKQSGAGQYKVICLEPLRDEPTYKLEFSDET 574
Db 526 VDATVGRPRVNFRTTIQORABFDYLHKQSGAGQYGRVTVGYPRLRPGSKXEPEFENMI 585
Oy 575 FGSNIIPQFPYAVVKGFLDACKEKPLSGHLSGLRFLQDCAHMTDSNLSPTIRAGGA 634
Db 586 VGOAIPGSGFIPALKEGFEKANSGLLGHVPENRIRLYLTQASHAAYVSSSELAIFKAAIYA 645
Oy 635 LKQALNAATCIIIEPIIAVEVVAVNEFGQVYAGINRRHGVTITQDCEGVEDFTLYADVPL 694
Db 646 FRLCYTAARPIYLEBVMVLVELKVPTEFGYVYAGDINRRKGIIVGNDQBGDSDVTANVPL 705
Oy 695 NDMPGYSTELRSCTEBGGEYTMETYSRQCPCLPSTQGEVINYKY 736
Db 706 NNMFGYSTLSRMTQGBETMEYKESAVSNEVQAOVLNAVY 747

```

## RESULT 5

P84885  
hypothetical protein At2g45030 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: P84885  
R./Lin, X.; Kall, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, A.; Vanakken, S.B.; Umayam, L.; Tallon, L.;  
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A./Reference number: A84420; MIMD:20083487; PMID:10617197  
A/Accession: P84885  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-754 <STO>  
A/Cross-references: GB:AE002093; NID:94895248; PIDN:AMD32833.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g45030  
A/Map position: 2  
C/Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 50.5%; Score 1970; DB 2; Length 754;  
Best Local Similarity 54.8%; Pred. No. 1.4e-120; Indels 12; Gaps 5;  
Matches 385; Conservative 117; Mismatches 188;

35 MSSSGVYIPEKINIGISAHIDSGKTLTERVLYTGRIAKMEVKKGKGVGAMDSMEL 94  
58 WKES---MDKANIGISAHIDSGKTLTERVLYTGRIRHIEHVGROGVGAKMDSMDL 113  
95 ERKRGITIGSAAFTMMKQVNIITDPGHVDTTIEVERALRLDGAIVLCAVGVQCC 154  
114 ERKRGITIGSAAFTMMKQVNIITDPGHVDTTIEVERALRLDGAIVLCAVGVQCC 173  
155 TMVTVNOMKRYNVPFLTFIKLDRMGSPARALQOMRSKLNHTAFMOI PMEGNFKGI 214  
174 STIVDOMKRYNVPFLTFIKLDRMGSPARALQOMRSKLNHTAFMOI PMEGNFKGI 233  
215 VDLIERAIFYDDPSQIVRYGELPAELRAAATDHRQELIECVANSDEQIGEMFLEKIP 274  
234 IDLIHVAAYFFHSGSGENVVAGDI PADMEGLVGDKRELIEYSEVDVLAKEFLNDEPV 293  
275 STSDLKLAIRATLKSGFTFVFGSALKKGVPLLDVALEIYIPNSEVONTAII LKKCD 334  
294 SAELEBARATIRATIAKFPVFGSAFKKNGVOPLLDGVVSPLEPSENVNVAL---DQ 349  
335 SKETKILMNSSHNHSPFVGLAFPLEVGRFGQLTVRSYQELKKGGDTIYNRTKKYR 394  
350 NNNEKRYTLTGSPDG--PLVALAFKLEBGRFGQLTVRYEGVYKAGDFINVTKRIR 407  
395 LORLARHADMEASTEEVYAGDICALFGIDCASGDTFTDKANSGLMESIHVPDVISI 454  
408 VPLVVMHNSDME-DIOEAHAGQIVAVFGIECASGDTFTD-GSVKVTMTSMNVPEVMSL 465  
455 AMKPSKNKDLKSKIGRTREDPTFKYFTDTENKETYISGMEGLHLIYAOQLREVG 514  
466 AVQPVSKDSGGQSKALNRFQKEDPTFRVGLDDESQOTIISGMEGLHLIYVERMRKRY 525  
515 CPCITGPKYAFRETTAPVPFDTKHKSGGAGQYGVYGLVEPLDPEDTYLTESDTE 574  
526 VDATVQKPRVNEFTTQRAEFYHLHKSGGAGQYGVYGLVEPLDPEDTYLTESDTE 585  
575 FGSNIPKQFVPAVEKFLDACEKPLSGHLSGLRVLQDGAHMTVDSNEISFIRAGEA 634  
586 VGGATISGFIPALIEKFKKAAKNSGLIGHVENLRIVLTDGASHAVDSSSLAKMAIYA 645  
635 LKQALANATLICLEPTMAVEVAPNEQGVYAGINRRHGVITGQGVEDYFTLVADVL 694  
646 FRLCYTAAPVILIEPMLVELKPTFEQGVYAGINRRHGVITGQGVEDYFTLVADVL 705  
695 NDMFGSTELRSCTEKSGEYMEYSRYQCLPSTOEDVINKY 736  
706 NMFQISTLSRSTQKGEFTMEYKSHAVSNVEVQALVNAY 747

RESULT 6  
T50308  
probable translation elongation factor EF-G SPBC1306.01c precursor, mitochondrial [setml1  
C/Species: Schizosaccharomyces pombe  
C/Date: 09-Jun-2000 #sequence\_revision 03-Nov-2000 #text\_change 03-Nov-2000  
C/Accession: T50308; T40449  
R./Brown, S.; Harrie, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, November 1999  
A./Reference number: 225060  
A/Accession: T50308  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-252 <BR>  
A/Cross-references: EMBL:AL13303; PIDN:CAB62091.1; GSPDB:GN00067; SPDB:SPBC1306.01c  
A/Experimental source: strain 972h(-); cosmid c1306  
R./Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M.  
submitted to the EMBL Data Library, August 1999  
A./Reference number: 221929  
A/Accession: T40449  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 213-770 <LN>  
A/Cross-references: EMBL:AL109822; PIDN:CAB52624.1; GSPDB:GN00067; SPDB:SPBC409.22c  
A/Experimental source: strain 972h(-); cosmid c409  
C/Keywords: mitochondrial  
F/60-208/Domain: translation elongation factor Tu homolog <ETU>

Query Match 48.8%; Score 1903.5; DB 2; Length 770;  
Best Local Similarity 54.4%; Pred. No. 3.2e-116; Indels 33; Gaps 10;  
Matches 390; Conservative 111; Mismatches 183;

44 EKIRNIGISAHIDSGKTLTERVLYTGRIAKMEVKKGKGVGAMDSMELERKGIITQ 103  
57 KQIRNIGISAHIDSGKTLTERVLYTGRIRKIDHEVRKQDNGAMDMELEREKGIITQ 116  
104 SAATTTM-----KQVNIITDPGHVDTTIEVERALRLDGAIVL 145  
117 SAATTCERTVTDQIEANEKQTDPEKSYNINIITDPGHVDTTIEVERALRLDGAIVL 176  
146 CAVGVQCCOTVTVNOMKRYNVPFLTFIKLDRMGSPARALQOMRSKLNHTAFMOI PM 205  
177 CAVGVQCCOTVTVNOMKRYNVPFLTFIKLDRMGSPARALQOMRSKLNHTAFMOI PM 236  
206 GLEGNFKGIVDLIERAIFYDDPSQIVRYG-ELPAELRAAATDHRQELIECVANSDEQL 264  
237 GQEDLLEGVVDLIQMRALYNNKSGKEKIEISQVPENLIELAKERSALIELKADLDEI 296  
265 GEMFLEEKIPISDILKAIIRATLKSGFTFVFGSALKKGVPLLDVALEIYIPNSEVQ 324  
297 ADIYMEEDPTPEQIMGAIRRTTARKETPVLMGSAIENVQSVLDAVCYLLPMPSEVE 356  
335 NYAIIKKDDSEKTKILMNSSHNHSPFVGLAFPLEVGRFGQLTVRSYQELKKGGDTI 384  
357 NIA-LNAAD--SEKVSILPSS--EKPLVALAFLEBGRFGQLTVRYLIIYQGTLRKGYI 410  
385 YNTRTRKRVRLQRLARHADMEASTEEVYAGDICALFGIDCASGDTFTDKANSGLMES 444  
411 YVNVSTKIKIKYSLRVHNSDME-BIEKVEAGGICALFGIDCASGDTFTDSSVS-YMTS 468  
445 IHVPDVISIAMKPSKNKDLKSKIGRTREDPTFKYFTDTENKETYISGMEGLHLI 504  
469 MEVPPEVVISLILKPSK-DTTSFSKALNRFQKEDPTFRVGLDDESQOTIISGMEGLHL 527  
505 YAOQLREYGCPCITGPKYAFRETTAPVPFDTKHKSGGAGQYGVYGLVEPLDPE 563  
528 YVERRRREYKQDCETGERVARETTLSKVPSPSYLHKQSGGAGQYARVGEYIEYMDGVE 587



Db 358 N---NTEDEIFEGSGGIYALFGEICASGDTFCD-PSINYSMTSMEFIPPVISLISVYKPD 412  
Qy 461 KNDLEKSKGIGRFTBEDPTEFKYFPTENKETVISCAGELHLEIYQORLEREGCCICG 520  
Db 413 KKSADMAKALGFTFEDPTEFKYVDESNETIIOGMBLHEVIERMKREKAEVETG 472  
Qy 521 KPRVARETITTAVPDPFTHKKQSGAGQYGVYIGVLEPLDPEDYKLEFSDETFGSNIP 580  
Db 473 MPQVARETITRAKEFNTHKKQSGAGQGVAVGMEPLDKKGET-YEFVNLKGGVLP 531  
Qy 581 KQVPAVEKGFLLDACEKGPLSGHKSGLRVLQDGAHMDVSNESIFIRAGSALKQALA 640  
Db 532 TEYIPSCDKGFQKAMERGTLLGPIYDIKTINDGQYHIVDSSDIAFOLAIGAFAEAYE 591  
Qy 641 NATLCILEPIMAVEVVAAPNEFOGOVYAGINRRHGVITG--QDGVEDYFTLYADVPINDMF 698  
Db 592 KAKPTLLEPIMKTLLEGPTFEQGNMGLNQRGIIITGSLEDG--SFSKVEAEVPLSEMF 649  
Qy 699 GYSTELRSCTEGKGYTMEYSRYOPLPSTQEDVINK 735  
Db 650 GFSTVLRSSITQKAEFSMEFLRYGK-VPSITIFDELNK 685

## RESULT 9

S04429 translation elongation factor EF-G - *Synechococcus* sp. (strain PCC 6301)  
C:Species: *Synechococcus* sp.  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Jan-2001  
C:Accession: S04429  
R:Memg, B.Y.; Shinozaki, K.; Sugitara, M.  
Mol. Gen. Genet. 216, 25-30, 1989  
A>Title: Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-A: reference number: S04426; M01D:89281486; PMID:2499762  
A:Accession: S04429  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-694 <MEN>  
A:Cross-references: GB:X17442; NID:g11204; PIDN:CAJ3495.1; PID:g1405432  
A>Note: the source is designated as *Anacystis nidulans*  
C:Genetics:  
A:Gene: fus  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog  
C:Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis  
F:11-138/Domain: translation elongation factor Tu homology <ETU>  
F:17-24/Region: nucleotide-binding motif A (P-loop)  
F:135-138/Region: GTP-binding NKXD motif  
F:111-313/Region: GTP-binding SAK/L motif  
F:23,24,62,135,136,138,311/Binding site: Mg-GTP (lys, thr, thr, Asn, lys, Asp, Ser) #sta

Query Match 38.3%; Score 1493; DB 1; Length 694;  
Best Local Similarity 44.5%; Pred. No. 1.7e-89;  
Matches 315; Conservative 128; Mismatches 229; Indels 36; Gaps 9;  
Qy 41 IPNEKIRNIGISAHIDSCKTTLTERVLYTGRIRAKHEVKGKDGAVAMDSELERQGI 100  
Db 5 VPLEKVRNIGIAHIDAGKTTTERTILFYSGVHKIGEY-HDG-NAATDMWEOEKERGI 61  
Qy 101 TIOSAATPFMMKQVNIINIDTGHVDTTEVERALVLDGAVLVLCVAGVQCGMTVNR 160  
Db 62 TITAAALISTSWKQYRNIINIDTGHVDTTEVERSKMVLGAVAVFCSVGQVQGSTVNR 121  
Qy 161 QMKRVVPLTFINKLDRMGSNPARALQOMRSKLNNTAFMOIPMGLGKFGKIVLIEER 220  
Db 122 QADRVSVPRIVFNKMDRIGADFFKYYGQIRDRVRANAVPIQIPIGABSDFOGIVLVEM 181  
Qy 221 RAIVPDGDSQIVRYGEIPAELEAAATDHRQELIECVANSDEQGMFLBKIPISIDL 280  
Db 182 KAHITVNDLGTDLVIDIPAELOETAERSKVVEAVATDEALDKYFEDGDLSTEDIK 241  
Qy 281 LAIRATLKRS---FPFVPLGSLKRGVQPLLDATLEVLNPSVQ-----NVAL 328  
Db 242 AGURKVVLLIQGNRLVPMLCGSAFNKGVQLLDVAVELLPSPODIPPIQGTLPDGEVAL 301

Qy 339 LNKDSDSEKTKILMNSSRHSNHPFVGLAFLVEGRFGOLTYVRSYOGELKKGDTIYNTR 388  
Db 302 RPSDEA-----PESALAFKIMADPYGRILFVRYSGIILQKGSVYNAT 345  
Qy 389 TRKXVRLQRLAMADMMEAETEEVYADICAFGI-DCASGDTFTDKANGLSMESHV 447  
Db 346 KKKKVRERLLIILKAD-DRIEVDLRAGDLGVLGKLPFTGDTLDDQNP-IILSELP 403  
Qy 448 PDPVLSIMKPSKNDLEKFSKIGRFTREDPTEFKYEDTENKETVISGMBLHEIYAQ 507  
Db 404 PEPVLSAVERPTKNDMEKLSKALALSEBDPTFVSVSDSETNQVILGMBLHEIYD 463  
Qy 508 FLREYGCPCITGKPKVAFRETITTAVPDPFTHKKQSGAGQYGVYIGVLEPLDEYTK 567  
Db 464 RMLREYKVEANIGAPQVAVRETVRKAVAEKFFVYQSGKQYGVHIVLELPAERG--TG 521  
Qy 568 LEFSEPTGSGNIPKQFVPAVEKGFLLDACEKGPLSGHKSGLRVLQDGAHMDVSNESIF 627  
Db 522 FEFVSKIVGTVPKXYVGPABQGMKETSGLVAGYPLIDIKATLVDSYHDVDSSEKAF 581  
Qy 628 IRAGSALKQALANATLCILEPIMAVEVVAAPNEFOGOVYAGINRRHGVITGQDGVEDYFT 687  
Db 582 KTAGSMALKEAVRKADPVLEPVMKVEVEBEDPLGSMGLISRGQIBQATNGTAT 641  
Qy 688 LYADVPLNDMFGYSTELRSCTEGKGYTMEYSRYOPLPSTQEDVINK 735  
Db 642 VSAKVPFLAMEFYATDLDSMTQGRGIFMERSQYVEVPRVAERTIAK 689

## RESULT 10

G71847 translation elongation factor EF-G (ef-g) - *Helicobacter pylori* (strain J99)  
C:Species: *Helicobacter pylori*  
A:Variety: strain J99  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Feb-2001  
C:Accession: G71847  
R:Alim, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; M01D:99120557; PMID:9923682  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-692 <ARN>  
A:Cross-references: GB:AE001539; GB:AE001439; NID:g4155706; PIDN:AAD06689.1; PID:g415570  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: fusA  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:11-138/Domain: translation elongation factor Tu homology <ETU>  
F:17-24/Region: nucleotide-binding motif A (P-loop)  
F:135-138/Region: GTP-binding NKXD motif

Query Match 38.2%; Score 1489; DB 2; Length 692;  
Best Local Similarity 45.7%; Pred. No. 3.1e-89;  
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;  
Qy 42 IPNEKIRNIGISAHIDSCKTTLTERVLYTGRIRAKHEVKGKDGAVAMDSELERQGI 101  
Db 6 PLNIRNIRNIGIAHIDAGKTTTERTILFYTGSHKIGEY-HDG-AATDMWEOEKERGIT 62  
Qy 102 IOSAATPFMMKQVNIINIDTGHVDTTEVERALVLDGAVLVLCVAGVQCGMTVNRQ 161  
Db 63 ITSAAITCFMWDQIINLIDTGHVDTTEVERSKMVLGAVSVFCSVGQVQGSTVNRQ 122  
Qy 162 MKRVVPLTFINKLDRMGSNPARALQOMRSKLNNTAFMOIPMGLGKFGKIVLIEER 221  
Db 123 ANKIVSVPRIVFNKMDRIGANFVYVENQIKQRLKANPPPINIPIGABDTFGIVDLVQMK 182  
Qy 222 AIYFDGDSQIVRYG--EIPAELEAAATDHRQELIECVANSDEQGMFLBKIPISIDL 279



Db 183 ATWYNNNE-TWGAAYDVBEIIPSDLEKAKQREVLVAVAHFOEALMKEKLGGEELDEI 241  
Qy 280 KLAIRATLKRSEFTFVFLGSALEKNGVPLLDVALEVLNPNSEVQVATILNKKDSKEKT 339  
Db 242 KKGITKGCILMNSVPLVPLCGSSFKNGKVGTLDDAVIYLPAPTEVVDI----KGIDPKTEE 297  
Qy 340 KILMNSSRHNSHPFVLGAPLEVGARF-GQLTYVRSYQGBLKKGDTIYNTTRKKVRLQRL 398  
Db 298 EVFVKS--DDEFFAGLAFKIMTDPFVGQLTFFVRVYRGKLESGSYVNSTKCKEKVGR 355  
Qy 399 AMHADMMEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMSIHVPDPVISIMK 457  
Db 356 LKMSHNRRE-DIKEVYAGELCAVNGKLDLTGTDLCDENKA-VLERMEFPEPVHIAVE 413  
Qy 458 PSNKNDLEKFSKIGRFTREDPTFKYVPDTEKKEVVISGMEHLIYVAQRLEREYGCPC 517  
Db 414 PRTKADQEKMGVALGKLAEDEBDFRVMTOEBTGQTLIGMGEHLIYVRLKREKRYEA 473  
Qy 518 ITGKRVAPRETITAPVPDFTHKQSGAGQYGVYIGVLEPLDPEDYTKLEFSDETFGS 577  
Db 474 EIGQPVAFRETITRIRSSVSKHKYAKQSGRGQYGHVFIKLEPEKG--SGYEFVNBISGG 531  
Qy 578 NIPKQVPAVEKGFPLDAEKGKPLSGHKLGLRFVLODGAHMYDSNEISFIRAGGALKQ 637  
Db 532 VLPKEVYIPVNDGIOAMONGVLAGYPVVDKFTLVDGSHVDVDSSEMAFKIAGSVAFFE 591  
Qy 638 ALANATLILBEPIMAVEVAPNEFOGOVYAGINRRHGVITGQGVEDYFTLVADVPILNDM 697  
Db 592 ASRAANPVILBEMKVEVEVPEVEMDVIGDLNRRGQINSMDRLGLKIYVAFPLVEM 651  
Qy 698 FGYSTELRSCTEGKGEYMEYSRYQPCLPSTOEDVINK 735  
Db 652 FGYSTDLRSATQGRGTYSMEFPHYGEVPSNIAKEIYK 689

RESULT 11  
C4669  
translation elongation factor EF-G - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 02-Feb-2001  
C:Accession: C64669  
R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khakhria, H.G.; Glodok, A.; McKenna, J.D.; Keilley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. A:Reference number: A64520; MID:9739467; PMID:9252185  
A:Accession: C64669  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-692 <TOM>  
A:Cross-references: GB:AB000625; GB:AB000511; NID:g2314349; PIN:AA08239.1; PID:g231435  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:11-138/Domain: translation elongation factor Tu homolog <ETU>  
F:17-24/Region: nucleotide-binding motif A (P-loop)  
F:135-138/Region: GTP-binding NKXD motif

Query Match 38.1%; Score 1487; DB 2; Length 692;  
Best Local Similarity 45.7%; Pred. No. 4.2e-89;  
Matches 319; Conservative 129; Mismatches 232; Indels 18; Gaps 11;  
Qy 42 PNEKIRNGISAHIDSGKTLTERVLYYTGRIAKNHEVKGXGVAGVANDSMELEROGIT 101  
Db 6 PLNRIRNGIAAHIDGKTTTERILFYGVSHKIGEV--HGG-AATMDMMQEKERGIT 62  
Qy 102 IOSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCANGVOCOTMTVNRQ 161  
Db 63 ITSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCANGVOCOTMTVNRQ 122  
Qy 162 MKRINVPFLFTINKLDRMGSNPARALQQRSKLNHNTAFVQIPMGLENGKGIVDLIEER 221

Db 123 ANKYGPRVVFVKNKDRIGANFYVNEQIKRLKANPVPINPIGAEDTFLFIVLQWAK 182  
Qy 222 AIFPGDQSQIYRYG--EIPAEIRAAATDHRRELECVANSSEOLGEMPLEKISISL 279  
Db 183 ATWYNNNE-TWGAAYDVBEIIPSDLEKAKYREKLEVAHFOEALMKEKLGGEELDEI 241  
Qy 280 KLAIRATLKRSEFTFVFLGSALEKNGVPLLDVALEVLNPNSEVQVATILNKKDSKEKT 339  
Db 242 KKGITKGCILMNSVPLVPLCGSSFKNGKVGTLDDAVIYLPAPTEVVDI----KGIDPKTEE 297  
Qy 340 KILMNSSRHNSHPFVLGAPLEVGARF-GQLTYVRSYQGBLKKGDTIYNTTRKKVRLQRL 398  
Db 298 EVFVKS--DDEFFAGLAFKIMTDPFVGQLTFFVRVYRGKLESGSYVNSTKCKEKVGR 355  
Qy 399 AMHADMMEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMSIHVPDPVISIMK 457  
Db 356 LKMSHNRRE-DIKEVYAGELCAVNGKLDLTGTDLCDENKA-VLERMEFPEPVHIAVE 413  
Qy 458 PSNKNDLEKFSKIGRFTREDPTFKYVPDTEKKEVVISGMEHLIYVAQRLEREYGCPC 517  
Db 414 PRTKADQEKMGVALGKLAEDEBDFRVMTOEBTGQTLIGMGEHLIYVRLKREKRYEA 473  
Qy 518 ITGKRVAPRETITAPVPDFTHKQSGAGQYGVYIGVLEPLDPEDYTKLEFSDETFGS 577  
Db 474 EIGQPVAFRETITRIRSSVSKHKYAKQSGRGQYGHVFIKLEPEKG--SGYEFVNBISGG 531  
Qy 578 NIPKQVPAVEKGFPLDAEKGKPLSGHKLGLRFVLODGAHMYDSNEISFIRAGGALKQ 637  
Db 532 VLPKEVYIPVNDGIOAMONGVLAGYPVVDKFTLVDGSHVDVDSSEMAFKIAGSVAFFE 591  
Qy 638 ALANATLILBEPIMAVEVAPNEFOGOVYAGINRRHGVITGQGVEDYFTLVADVPILNDM 697  
Db 592 ASRAANPVILBEMKVEVEVPEVEMDVIGDLNRRGQINSMDRLGLKIYVAFPLVEM 651  
Qy 698 FGYSTELRSCTEGKGEYMEYSRYQPCLPSTOEDVINK 735  
Db 652 FGYSTDLRSATQGRGTYSMEFPHYGEVPSNIAKEIYK 689

RESULT 12  
AC2348  
translation elongation factor EF-G [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC2348  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Watanabe, A.; Iriuguchi, R.; Kaneko, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Anabaena. A:Reference number: A81807; MID:21595285; PMID:11759840  
A:Accession: AC2348  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-692 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA076037.1; PID:g17133474; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: fus  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 37.8%; Score 1472; DB 2; Length 692;  
Best Local Similarity 44.9%; Pred. No. 4e-88;  
Matches 319; Conservative 124; Mismatches 223; Indels 44; Gaps 11;  
Qy 42 PNEKIRNGISAHIDSGKTLTERVLYYTGRIAKNHEVKGXGVAGVANDSMELEROGIT 101  
Db 6 PLEKIRNGIAAHIDGKTTTERILFYSGIHHKIGEV--HEGT-AVIDMDQERERGIT 62  
Qy 102 IOSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCANGVOCOTMTVNRQ 161  
Db 63 ITAAISTSMQDVNINIITDPGHVDFTEVERALRVLDGAVLVCANGVOCOTMTVNRQ 122



```

Db      63  ITSAATTCFWMKQHOINLIDTGHVDFTEIEVERSMRLDGAIVAFCSVGVQOQSESTVMQ 122
Qy      162  MKRYVNPFLTFINKLDMGNSPAPALQOMRSKLNNHTAFMOIPMGLEGFKIVLIEBR 221
Db      123  ANKGVPRVVFVNKNDRIQANFYNNEDQIRNLKANPVPIQIPIGHEDNFKGVITDLYWK 182
Qy      222  AIYFDGDSQIVRY--GEIPAEIARAATDHOELIECVANSDEQJEMLEEKIPSIDL 279
Db      133  ALWVEDD--TKPTDYVEKEIPAEIKKAEYRTKMEIAVSEITSDLMEXKLGSEELSLBEI 241
Qy      280  KLAIRATLKASFTFVFLGSAIKNKGVQPLDVALELPLNPSBVQNYAILNKDDESKET 339
Db      242  KGIYAGCLSLISIVMLCCTAFKNKGVQPLDVAVAYLPAPEEVANI-----KGEYEDT 296
Qy      340  KILMNSRSHNSHPFVGLAFPLEVGR--GOLTYVRSYOGELKKGDITYNTRTKKYRLQRL 398
Db      297  EVSVKST--DGEFAGLAKKIMTDFVQGLTFVRVYRGCLSGSYAVANSTKCKEIKIGL 354
Qy      399  AMHADMMMEASTEVEVYAGDICALFGI--DCASGDTFTDKANSGLSMESIHPDPVSIAMK 457
Db      355  LKHGNNKHE--EIKVLYAGEIAGVVGKDTLTGDTLASBKIK--VILERMDFPDPVSIANE 412
Qy      468  PSNKNDLEKFSKIGRFTREDPTFKYYPFTENKEVTYISGMGLHLEIYAQRLEREYGCPC 517
Db      443  PCTKADQEKMSIALNKLAQEDPSFRVSTDESGQTIISGMGLHLEIIVDRMLRERKVA 472
Qy      518  ITGKRVAPRETTITAPVPFDFTHKKOSGAGQYGVKIVGLLEPLDEPDYTKLEFSDTEFS 577
Db      473  EYGOQVAVARETIRKRTVEQYRYAKOSGGRGQYGHFRLPLEBEG--SGYFVNDIKGG 530
Qy      578  NIPKQFVPAVEKGFPLDACEKGPLSGHKLSGRLFVLQDGAHMHVDSNEISFIRAGECALQ 637
Db      531  VIPKTYIPAVDGVGOEALONGVLAGPVVEDVATVYDGSYHEVDSEMAFKLAASMGFE 590
Qy      638  ALAANATLCLEPIYMAVEVVAENPEFQGVYIAGINRRHGVTITGQDVEDYFTLYADVPLNDM 697
Db      591  GARKGAVVLEBPMKVEVETPEDYMGDVIQDLNKRGVQNSMDERGNKXITIAFCPLAEM 650
Qy      698  FGSTELRSCTEGKGEYTMESRYOCPCLPSTOEDVINK 735
Db      651  FGYSTDLRSQTKGRATYSMEFPHYDEVPRNADEIITK 688

```

## RESULT 15

S04390

translation elongation factor EF-G - *Spirulina platensis*C1Species: *Spirulina platensis*

C1Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 02-Feb-2001

C1Accession: S04390

R1Buttarelli, F.R.; Calogero, R.A.; Tiboni, O.; Gualerzi, C.O.; Pon, C.L.

Mol. Gen. Genet. 217, 97-104, 1989

A1Title: Characterization of the *str* operon genes from *Spirulina platensis* and their evolution

A1Reference number: S04388; UID:89364697; PMID:2505055

A1Accession: S04390

A1Molecule type: DNA

A1Residues: 1-697 &lt;BUT&gt;

A1Cross-references: EMBL:X15646; NID:g47447; PIDN:CA33672.1; PID:g581727

C1Genetics:

A1Gene: *tus*

A1Start codon: GTG

C1Superfamily: translation elongation factor G; translation elongation factor Tu homolog

C1Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis

F.11-138/Domain: translation elongation factor Tu homology &lt;ETU&gt;

F.17-24/Region: nucleotide-binding motif A (P-loop)

F.113-138/Region: GTP-binding NKD motif

F.313-315/Region: GTP-binding SAK/L motif

F.23,24,62,135,138,139,313/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #acc

Query Match 37.3%; Score 1454; DB 2; Length 697;

Beet Local Similarity 44.0%; Pred. No. 6e-877;

Matches 309; Conservative 138; Mismatches 234; Indels 22; Gaps 11;

```

Qy      41  IPNEKIRNIGISAHIDSGKTLTERVLYYTGRIAMHEVKKDGVAVMDSMEIERORGI 100
Db      5  IPLERVNIGIAAHIDAKKTITTEILIFYSGVHMGEV--HEGT-AVTDMWAQERERGI 61
Qy      101  TIQSAATTTMKDNNINIIDTGHVDFTEIEVERBALRVLDGAVLYLCANVGVOQOMTYNR 160
Db      62  TITAAISTSWLDHINIIDTPGHVDFTEIEVERSRVLDGVIAVCSVGVQOQSESTVWR 121
Qy      161  OMKRYVNPFLTFINKLDMGNSPAPALQOMRSKLNNHTAFMOIPMGLEGFKIVLIEBR 220
Db      122  QAEYQVRIAPFINKMDRTGADFFKVVQIQRDLRAANAVPIQVPRGESDTHGLVDVAM 181
Qy      221  AIYFDGDSQIVRYG--EIPAEIARAATDHOELIECVANSDEQJEMLEB---KIPS 275
Db      132  KTYLYTNDLGTVDIQSDEIPREVQDLVAEYREKLEIAVAETDEALMEKYLEQLEGBALT 241
Qy      276  ISDLTAIRATLKRSFTFVFLGSAIKNKGVQPLDVALELPLNPSBVQNYAILNKDODS 335
Db      242  BEEIRHSLROGTIKGLIVPVICGSFKNRGVQRLDVAVDYLPAPTEVP--PIKGLPBG 299
Qy      336  KEKTKILMNSRSHNSHPFVGLAFPLEVGRFGOLTYVRSYOGELKKGDITYNTRTKKYRL 395
Db      300  EBGVRYADDDA-----PLSALAFAKMDPGRILTFVRVYSGVLQKSGYITNATKAKKARI 354
Qy      396  QRLAMHADMMMEASTEVEVYAGDICALFGI--DCASGDTFTDKANSGLSMESIHPDPVSI 454
Db      355  SRLIYAKSD--EIRIEVEIRAGDLGALGLKDTLTGDTICDEANS--IILESIYIPPVISV 412
Qy      455  AMKPSKNDLEKFSKIGRFTREDPTFKYYPFTENKEVTYISGMGLHLEIYAQRLEREYGC 514
Db      443  AVEPKTKDMKELSKALQSLSEEDPTFRVSIIDSEINQTVIAGMGLHLEIIVDRMLRERK 472
Qy      515  CPCIKGPKVAPRETTITAPVPFDFTHKKOSGAGQYGVKIVGLLEPLDEPDYTKLEFSDT 574
Db      473  VEANIGAOVAVARETIRKRTVEQYRYAKOSGGRGQYGHVLEBEGEG--SGPFVSKI 530
Qy      575  FGSNIPKQFVPAVEKGFPLDACEKGPLSGHKLSGRLFVLQDGAHMHVDSNEISFIRAGECA 634
Db      531  VGSVPKTYIPAVDGVGOEALONGVLAGPVVEDVATVYDGSYHEVDSEMAFKIAGSMA 590
Qy      635  LKQALANATLCLEPIYMAVEVVAENPEFQGVYIAGINRRHGVTITGQDVEDYFTLYADV 692
Db      591  IKNGVTKASPVLEBPMKVEVETPEDYMGDVIQDLNKRGVQNSMDERGNKXITIAFCPLAEM 650
Qy      693  PLNDMFGSTELRSCTEGKGEYTMESRYOCPCLPSTOEDVINK 735
Db      651  PLATMFGYATIDRSKTKGRGVFSMEFPHYDEVPRNSVAETITAK 693

```

Search completed: July 14, 2003, 18:23:19

Job time : 19.9366 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 48.2242 Seconds  
(without alignments)  
4965.877 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCLQSTRVWLRENGCHFPST.....MIVKKRYSTRSASSQSSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10730	100.0	2057	US-09-815-379-10	Sequence 10, Appl
2	10628	99.0	2058	US-09-815-379-17	Sequence 17, Appl
3	10247	95.5	2048	US-09-815-379-12	Sequence 12, Appl
4	5385	50.2	1031	US-09-764-868-647	Sequence 647, Appl
5	1789	16.7	348	US-09-764-868-1073	Sequence 1073, Appl
6	1789	16.7	348	US-09-955-999-101	Sequence 101, Appl
7	1492	13.9	2548	US-09-851-682A-1	Sequence 1, Appl
8	1461	13.6	1855	US-10-177-293-315	Sequence 315, Appl
9	1391	13.0	1285	US-10-205-823-273	Sequence 273, Appl
10	1391	13.0	1285	US-10-177-293-317	Sequence 317, Appl
11	1360	12.7	1016	US-10-044-303-2	Sequence 2, Appl
12	1353	12.4	1938	US-10-171-311-164	Sequence 164, Appl
13	1333	12.4	1972	US-10-171-311-162	Sequence 162, Appl
14	1329	12.4	765	US-10-044-303-1	Sequence 1, Appl
15	1328.5	12.4	1945	US-09-927-597-2	Sequence 2, Appl
16	1328.5	12.4	1979	US-09-927-597-4	Sequence 4, Appl
17	1327.5	12.4	2405	US-10-128-714-8290	Sequence 8290, Appl
18	1282	11.9	2099	US-10-128-714-3290	Sequence 3290, Appl
19	1206	11.2	845	US-09-927-597-14	Sequence 14, Appl

20	1201.5	11.2	849	9	US-09-927-597-10	Sequence 10, Appl
21	1201.5	11.2	852	9	US-09-927-597-12	Sequence 12, Appl
22	1114	10.4	772	9	US-09-927-597-8	Sequence 8, Appl
23	1012	9.4	699	9	US-09-927-597-6	Sequence 6, Appl
24	779	7.3	151	10	US-09-864-761-46352	Sequence 46352, A
25	779	7.3	151	10	US-09-864-761-48367	Sequence 48367, A
26	721	6.7	569	9	US-09-925-300-1583	Sequence 1583, Ap
27	713	6.6	544	9	US-10-106-698-5864	Sequence 5864, Ap
28	632	5.9	2568	10	US-09-866-108-3	Sequence 3, Appl
29	628	5.9	136	9	US-10-106-698-6839	Sequence 6839, Appl
30	549	5.1	570	10	US-09-764-853-551	Sequence 551, Appl
31	520	4.8	104	10	US-09-864-761-46427	Sequence 46427, A
32	487	4.5	94	10	US-09-864-761-48393	Sequence 48393, A
33	477	4.4	113	9	US-10-106-698-6068	Sequence 6068, Ap
34	388	3.6	80	10	US-09-864-761-46319	Sequence 46319, A
35	343	3.2	66	10	US-09-864-761-42111	Sequence 42111, A
36	313.5	2.9	245	10	US-09-925-302-686	Sequence 686, Appl
37	292.5	2.7	294	9	US-10-106-698-5501	Sequence 5501, Appl
38	288.5	2.7	2125	10	US-09-919-172-29	Sequence 29, Appl
39	264	2.5	53	9	US-09-764-891-2868	Sequence 2868, Appl
40	260	2.4	1239	9	US-09-291-417-13	Sequence 13, Appl
41	244.5	2.3	1233	9	US-09-291-417-89	Sequence 89, Appl
42	229.5	2.1	1175	10	US-09-771-161A-224	Sequence 224, Appl
43	229.5	2.1	1175	10	US-09-771-161A-225	Sequence 225, Appl
44	229.5	2.1	1175	10	US-09-771-161A-226	Sequence 226, Appl
45	228	2.1	240	9	US-10-101-487-75	Sequence 75, Appl

## ALIGNMENTS

RESULT 1									
US-09-815-379-10									
; Sequence 10, Application US/09815379									
; Publication No. US20030073613A1									
; GENERAL INFORMATION:									
; APPLICANT: RASTELLI, LUCA									
; TITLE OF INVENTION: ANTIGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS									
; TITLE OF INVENTION: ENCODING THE SAME									
; FILE REFERENCE: 10716/35									
; CURRENT APPLICATION NUMBER: US/09/815,379									
; CURRENT FILING DATE: 2001-03-22									
; PRIOR APPLICATION NUMBER: 60/191,134									
; PRIOR FILING DATE: 2000-03-22									
; NUMBER OF SEQ ID NOS: 17									
; SOFTWARE: Patentin Ver. 2.1									
; SEQ ID NO 10									
; LENGTH: 2057									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-09-815-379-10									
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2057; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	FCLQSTRVWLRENGCHFPSTVNSCAEGIVPRTDYGQFTYQSTTTHOKTANMPPTNE	60						
DB	1	FCLQSTRVWLRENGCHFPSTVNSCAEGIVPRTDYGQFTYQSTTTHOKTANMPPTNE	60						
QY	61	GVDDNASITELHGSIMTNLFORYKRNQIWTYIGSILASVNPYOPJAGLYBATMEQYR	120						
DB	61	GVDDNASITELHGSIMTNLFORYKRNQIWTYIGSILASVNPYOPJAGLYBATMEQYR	120						
QY	121	RHLGELPHIIPAIANEYCRLMKRDNOICILKSGSAGKTESTKLLIKFLSVISQOSLE	180						
DB	121	RHLGELPHIIPAIANEYCRLMKRDNOICILKSGSAGKTESTKLLIKFLSVISQOSLE	180						
QY	181	LSLKRTSCVEKALIESPIMEAFGNATTVYNNSSRGKPYQVNI CQKGNIQGRIVDC	240						
DB	181	LSLKRTSCVEKALIESPIMEAFGNATTVYNNSSRGKPYQVNI CQKGNIQGRIVDC	240						

QY 241 ILSQNRVVRQNGERNYHIFVALLAGLEBHEEREFEYLSTPENYHYLNGSCVEDKTIISD 300  
 DB 241 ILSQNRVVRQNGERNYHIFVALLAGLEBHEEREFEYLSTPENYHYLNGSCVEDKTIISD 300  
 QY 301 QSFREVITAMDWQSKKEEVRVSLRLLAGILHGNIEFTTAGAQSFTALGRABELL 360  
 DB 301 QSFREVITAMDWQSKKEEVRVSLRLLAGILHGNIEFTTAGAQSFTALGRABELL 360  
 QY 361 GLDPTQUTDALTORSMFLRGEELITPLANOQAVDSRSLAMALYACCFEVIKIKISRIK 420  
 DB 361 GLDPTQUTDALTORSMFLRGEELITPLANOQAVDSRSLAMALYACCFEVIKIKISRIK 420  
 QY 421 GNEDFKSIIGLIDIFGFENFEVNHFEQFNINAYANEKLOEYFNKHI FSLQOLEYREGIYWE 480  
 DB 421 GNEDFKSIIGLIDIFGFENFEVNHFEQFNINAYANEKLOEYFNKHI FSLQOLEYREGIYWE 480  
 QY 481 DITWINGECLDILEKKLGILALINEBHPQATDSTLEKLSOHANNHFYKPRVAVN 540  
 DB 481 DITWINGECLDILEKKLGILALINEBHPQATDSTLEKLSOHANNHFYKPRVAVN 540  
 QY 541 NFGVKHAYAGEVOYDVARGILEKNDTRFDLNLRSRPFYIDLFEHVSRRNODTLKC 600  
 DB 541 NFGVKHAYAGEVOYDVARGILEKNDTRFDLNLRSRPFYIDLFEHVSRRNODTLKC 600  
 QY 601 GSKHRRPTVSSQCFKVDLSHSLMATLSSNPFVRCIKPMNQKMPDQAVLVLNQLRYSG 660  
 DB 601 GSKHRRPTVSSQCFKVDLSHSLMATLSSNPFVRCIKPMNQKMPDQAVLVLNQLRYSG 660  
 QY 661 METETVIRKAGVAVRRPFQDFYKRYKVLNMLALPBDVAGKCTSLQLVDASNSEWQOLK 720  
 DB 661 METETVIRKAGVAVRRPFQDFYKRYKVLNMLALPBDVAGKCTSLQLVDASNSEWQOLK 720  
 QY 721 TKVFLRESLEQLEKREBEVSHAAMVIRAHVGLFAPLAKOYKVLVCVYIIONYAPFL 780  
 DB 721 TKVFLRESLEQLEKREBEVSHAAMVIRAHVGLFAPLAKOYKVLVCVYIIONYAPFL 780  
 QY 781 RRRFLHLKKAIVFOKQOLRQIARVYRQLLAKEBOEKKOEBEKKKREBERERER 840  
 DB 781 RRRFLHLKKAIVFOKQOLRQIARVYRQLLAKEBOEKKOEBEKKKREBERERER 840  
 QY 841 ERREALRAQOEETKQOELALQSKOEALITRELEKQENKQOEBELIRLEKEIEDIO 900  
 DB 841 ERREALRAQOEETKQOELALQSKOEALITRELEKQENKQOEBELIRLEKEIEDIO 900  
 QY 901 RMEQOELSTEASLOKLOERRDOELRLLEBACRAOEFLESLNDEIDECVRNIERSI 960  
 DB 901 RMEQOELSTEASLOKLOERRDOELRLLEBACRAOEFLESLNDEIDECVRNIERSI 960  
 QY 961 SGGSEFSSSILASACEKPNFNSQPYPEEVEDEGEADDAFKDSPNPSEHGSDQRTS 1020  
 DB 961 SGGSEFSSSILASACEKPNFNSQPYPEEVEDEGEADDAFKDSPNPSEHGSDQRTS 1020  
 QY 1021 GRTSDSSSEEDPYMNDTVVPTSPADSTVLAPSVQDSGLNSSGSESTYCMPONAD 1080  
 DB 1021 GRTSDSSSEEDPYMNDTVVPTSPADSTVLAPSVQDSGLNSSGSESTYCMPONAD 1080  
 QY 1081 LPSPDGDYDODDYEDGATTSQSSYTFNSYSQMSPTVRCVGTYSNGAVRFSSEGA 1140  
 DB 1081 LPSPDGDYDODDYEDGATTSQSSYTFNSYSQMSPTVRCVGTYSNGAVRFSSEGA 1140  
 QY 1141 QSSFEDESEDFSRFTDDELSTYRDSVYSCTVLPYFHSFLYMKGGLMNSMKRMCVLD 1200  
 DB 1141 QSSFEDESEDFSRFTDDELSTYRDSVYSCTVLPYFHSFLYMKGGLMNSMKRMCVLD 1200  
 QY 1201 EFTFLMRSQOALKQMLHKKGGSSSTLSRRNMKKMFWLRQSKLMYFENDSEKLGTV 1260  
 DB 1201 EFTFLMRSQOALKQMLHKKGGSSSTLSRRNMKKMFWLRQSKLMYFENDSEKLGTV 1260  
 QY 1261 EYRTAKEIINDNTKENGIDITIMADRTFHLIASPEDASQWFSVLSOVHASTDOEIOMPD 1320  
 DB 1261 EYRTAKEIINDNTKENGIDITIMADRTFHLIASPEDASQWFSVLSOVHASTDOEIOMPD 1320  
 QY 1321 EQANPONAVGTLVDGLIDSVCASDSPDRPNSFVITANRVLMCNADPEEMHMTTLQOR 1380

DB 1321 EQANPONAVGTLVDGLIDSVCASDSPDRPNSFVITANRVLMCNADPEEMHMTTLQOR 1380  
 QY 1381 SKGDTRVEGOEFTVGMHLKEVKNSPKMSLKLKRWPLTNSLDYKSSSKNMLKGT 1440  
 DB 1381 SKGDTRVEGOEFTVGMHLKEVKNSPKMSLKLKRWPLTNSLDYKSSSKNMLKGT 1440  
 QY 1441 LVNLISCSVVPDEKIFKSTGYMNTVYGRKHCYRLYTLGLNEATRWSSVIONVDTKAP 1500  
 DB 1441 LVNLISCSVVPDEKIFKSTGYMNTVYGRKHCYRLYTLGLNEATRWSSVIONVDTKAP 1500  
 QY 1501 IDTPTQOLIODIKENCLNSDVVEQIYKKNPILRYTHHPHSLPLLPYGDINLNLKXDG 1560  
 DB 1501 IDTPTQOLIODIKENCLNSDVVEQIYKKNPILRYTHHPHSLPLLPYGDINLNLKXDG 1560  
 QY 1561 YTTLODEAKITNSLOQESMSDPIPIQGLIQTHDLRPLDELYCOLIKQTNKVPHPG 1620  
 DB 1561 YTTLODEAKITNSLOQESMSDPIPIQGLIQTHDLRPLDELYCOLIKQTNKVPHPG 1620  
 QY 1621 SVGNLYSQILTCLSTFLPSRGIKYLKFLHKLREOPGTMEKALFTYESLKTTC 1680  
 DB 1621 SVGNLYSQILTCLSTFLPSRGIKYLKFLHKLREOPGTMEKALFTYESLKTTC 1680  
 QY 1681 REFVPSRDEIHALHROEWTSTVYCHGGGCKITINSHTAGEVVEKLRGLAMEDSRNM 1740  
 DB 1681 REFVPSRDEIHALHROEWTSTVYCHGGGCKITINSHTAGEVVEKLRGLAMEDSRNM 1740  
 QY 1741 PALFENGVHVDKISRITVAVDLAKFEKLATSEVGDPMKFFRYLCEFLTDNVPKOS 1800  
 DB 1741 PALFENGVHVDKISRITVAVDLAKFEKLATSEVGDPMKFFRYLCEFLTDNVPKOS 1800  
 QY 1801 VEFAPMEQAEHVAIGHHPAPEBNLOVLAALRYLOQDYTLTHAIPLEBEVYSIORLK 1860  
 DB 1801 VEFAPMEQAEHVAIGHHPAPEBNLOVLAALRYLOQDYTLTHAIPLEBEVYSIORLK 1860  
 QY 1861 ARIISOSTKTPPCERLEKRRISFLEGTLRSSRTSSVVRQKYEBOOMLDMITKEVYSAR 1920  
 DB 1861 ARIISOSTKTPPCERLEKRRISFLEGTLRSSRTSSVVRQKYEBOOMLDMITKEVYSAR 1920  
 QY 1921 ASIIDKMRFOGOMNOEOMAKYMALIKEMPGVSTLPDVCKEGGFPOEIMIGVSADAYS 1980  
 DB 1921 ASIIDKMRFOGOMNOEOMAKYMALIKEMPGVSTLPDVCKEGGFPOEIMIGVSADAYS 1980  
 QY 1981 YKKEGGRPLEVFOYEHLISFGAPLANTYKIIVDERELLFETSEVVDVAKLMKAYISMTV 2040  
 DB 1981 YKKEGGRPLEVFOYEHLISFGAPLANTYKIIVDERELLFETSEVVDVAKLMKAYISMTV 2040  
 QY 2041 KRRYSTTSASQSSR 2057  
 DB 2041 KRRYSTTSASQSSR 2057  
 RESULT 2  
 US-09-815-379-17  
 ; Sequence 17, Application US/09815379  
 ; Publication No. US20030073613A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RASTELLI, LUCA  
 ; APPLICANT: GERITSEN, MARY  
 ; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS  
 ; TITLE OF INVENTION: ENCODING THE SAME  
 ; FILE REFERENCE: 10716/35  
 ; CURRENT APPLICATION NUMBER: US/09/815,379  
 ; PRIOR FILING DATE: 2001-03-22  
 ; PRIOR APPLICATION NUMBER: 60/191,134  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 17  
 ; LENGTH: 2058  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-815-379-17

Query Match	99.0%; Score 10628; DB 9; Length 2058;	
Best Local Similarity	99.3%; Pred. No. 0;	
Matches 2043; Conservative	5; Mismatches 7; Indels 2; Gaps 2;	
Qy	1 FCLQGTWVLRNGGHFSTVNSCAEGIVFRTDVGQVFTYKQSTTHQKVTAMHPTNEE	60
Db	4 FETGEVRLRENGGHFSTVNSCAEGIVFRTDVGQVFTYKQSTTHQKVTAMHPTNEE	63
Qy	61 GVDMAASLTELHGSSIMYNLFOYRKNOITWTYIGSLASVNPQIPAGIYEPATMEQYR	120
Db	64 GVDMAASLTELHGSSIMYNLFOYRKNOITWTYIGSLASVNPQIPAGIYEPATMEQYR	123
Qy	121 RHAGELPHIIPAIANEVCYRCLWKRDHNOICILKGSAGICTESTKILKFLSVISQOSLE	180
Db	124 RHAGELPHIIPAIANEVCYRCLWKRDHNOICILKGSAGICTESTKILKFLSVISQOSLE	183
Qy	181 LSLKETSCEVERAIILESSPIMEAFGNAKTVYNNNSRFGKIVQANI CQKGNIOGGRIVDC	240
Db	184 LSLKETSCEVERAIILESSPIMEAFGNAKTVYNNNSRFGKIVQANI CQKGNIOGGRIVDY	243
Qy	241 ILSSONRVVRONGEENYHI FYALLAGLEHEERESEFYLSIPENYHILNOSGVEDKTISD	300
Db	244 IL-EXONRVVRONGEENYHI FYALLAGLEHEERESEFYLSIPENYHILNOSGVEDKTISD	302
Qy	301 QESFEVITAMVMOFSKEEVRVSRLLAGILHLGNIIEFTAGAQVSEKTAIGSAAELL	360
Db	303 QESFEVITAMVMOFSKEEVRVSRLLAGILHLGNIIEFTAGAQVSEKTAIGSAAELL	362
Qy	361 GLDPQLDALTQSRMFLRGEELTPLANVOAVDSRDSI LAMALYACCFEWTIKTINSRIK	420
Db	363 GLDPQLDALTQSRMFLRGEELTPLANVOAVDSRDSI LAMALYACCFEWTIKTINSRIK	422
Qy	421 GNEDEKSGILDI FGEENFVNHFEOFINVANEKLOEYFNKHIFSLDELEYSRSEGLWE	480
Db	423 GNEDEKSGILDI FGEENFVNHFEOFINVANEKLOEYFNKHIFSLDELEYSRSEGLWE	482
Qy	481 DIDWIDNGECLDLEIKKGLILALINEESHFPQATDSTLEKLSQHANNH FYVKPRVAVN	540
Db	483 DIDWIDNGECLDLEIKKGLILALINEESHFPQATDSTLEKLSQHANNH FYVKPRVAVN	542
Qy	541 NGVGHVAGEVOYDVRGILEKRDTRFDLNLALRESRFDYIDLEFHVSSRNNODTLKC	600
Db	543 NGVGHVAGEVOYDVRGILEKRDTRFDLNLALRESRFDYIDLEFHVSSRNNODTLKC	602
Qy	601 GSKHRRPTVSSQPKYDLSHSLMATTSSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Db	603 GSKHRRPTVSSQPK-DSLHSLMATTSSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	661
Qy	661 MLETVRIRKAGYAVRRPQDFYKRYKVLNRNLALPEDVSGKCTSLQLYDASNSEWQLGK	720
Db	662 MLETVRIRKAGYAVRRPQDFYKRYKVLNRNLALPEDVSGKCTSLQLYDASNSEWQLGK	721
Qy	721 TKVPLRESLEQLEKREBEVSHAAVIRAHVIGFLARIQYRKVLYCVVITQNTAFELL	780
Db	722 TKVPLRESLEQLEKREBEVSHAAVIRAHVIGFLARIQYRKVLYCVVITQNTAFELL	781
Qy	781 RRRFLHLKKAALVFOKOLAGIARARYROLALAEKREOEKKOESEKKKREBEERER	840
Db	782 RRRFLHLKKAALVFOKOLAGIARARYROLALAEKREOEKKOESEKKKREBEERER	841
Qy	841 EERRBAELRAQOEETFKQOELALQSKQEAELTRLEIKQENKQYAEILRLKEKETEDQ	900
Db	842 EERRBAELRAQOEETFKQOELALQSKQEAELTRLEIKQENKQYAEILRLKEKETEDQ	901
Qy	901 RKKEQOELSTLASLOKLOERRDQELRLLEBA CRAAQFLESANFDEIDECRYNIERSL	960
Db	902 RKKEQOELSTLASLOKLOERRDQELRLLEBA CRAAQFLESANFDEIDECRYNIERSL	961
Qy	961 SGGSEFSESLASACEKPNFNSOPYPREEVDEGEADDDAFKQSPNPSSEHGHSDQRTS	1020
Db	962 SGGSEFSESLASACEKPNFNSOPYPREEVDEGEADDDAFKQSPNPSSEHGHSDQRTS	1021

Qy	1021 GIRISDSSSEEDPYNDITVPTSPSADSTVLIAPSVODSGSLHNSSCGEISTY CMPQONAGD	1080
Db	1022 GIRISDSSSEEDPYNDITVPTSPSADSTVLIAPSVODSGSLHNSSCGEISTY CMPQONAGD	1081
Qy	1081 LPSPDGDYDODDVEDDAITSGSSVTFSNSGSGOWSPDYRCSVCTTYNSSGAYRPSSEBA	1140
Db	1082 LPSPDGDYDODDVEDDAITSGSSVTFSNSGSGOWSPDYRCSVCTTYNSSGAYRPSSEBA	1141
Qy	1141 QSSFEDSEEDPSRDTDELSTYRDSVYSCVTLTYFHSFLYMKGGLANMSWRRCRCVLD	1200
Db	1142 QSSFEDSEEDPSRDTDELSTYRDSVYSCVTLTYFHSFLYMKGGLANMSWRRCRCVLD	1201
Qy	1201 EFTFLFRSKQEBALKQGMHLKKGGSSTLSRRNKKRWFLRQSKLMEYENDSEBKLGCTV	1260
Db	1202 EFTFLFRSKQEBALKQGMHLKKGGSSTLSRRNKKRWFLRQSKLMEYENDSEBKLGCTV	1261
Qy	1261 EYRTAKEIINDNTTKENGIDITIMADTTFHLIASPEPDAQWFSVLQVHASTQOEIOEMND	1320
Db	1262 EYRTAKEIINDNTTKENGIDITIMADTTFHLIASPEPDAQWFSVLQVHASTQOEIOEMND	1321
Qy	1321 EONAPONAVGTLDVGLIDSVCASDSPDRPNSFVITANRVLHCNADTPBEMHMITLQOR	1380
Db	1322 EONAPONAVGTLDVGLIDSVCASDSPDRPNSFVITANRVLHCNADTPBEMHMITLQOR	1381
Qy	1381 SKGDRVREGOEPIVAGMLHKEVONS PKWSLKLKRWFLTNSLDYKXSSEKNAKLGT	1440
Db	1382 SKGDRVREGOEPIVAGMLHKEVONS PKWSLKLKRWFLTNSLDYKXSSEKNAKLGT	1441
Qy	1441 LVLNLSLSVPPDEKIFKETGYMNTYVGRKRCYLLTYKLNEARMSVIONVTDTKAP	1500
Db	1442 LVLNLSLSVPPDEKIFKETGYMNTYVGRKRCYLLTYKLNEARMSVIONVTDTKAP	1501
Qy	1501 IDTPTQOOLIODIKENCLNSDVVEQIYKKNPILRYTHHP LHSPLPLPYGDIMNLNLKDGK	1560
Db	1502 IDTPTQOOLIODIKENCLNSDVVEQIYKKNPILRYTHHP LHSPLPLPYGDIMNLNLKDGK	1561
Qy	1561 YTTLODEAIKIFNSLQOLESMSDPPIIIGILQTHDRLPLDELYCQILKQTNVPHRG	1620
Db	1562 YTTLODEAIKIFNSLQOLESMSDPPIIIGILQTHDRLPLDELYCQILKQTNVPHRG	1621
Qy	1621 SVGNLYSMQILTCLSTPLPSRGIILKYLKFLKRI REOPPGTMEBKRYALPTYESLKTTC	1680
Db	1622 SVGNLYSMQILTCLSTPLPSRGIILKYLKFLKRI REOPPGTMEBKRYALPTYESLKTTC	1681
Qy	1681 REFVPSRDEIEALIHROEMTSTVYCHGGGSCKITINSHTJAGEVEKILRGLAMEDSRM	1740
Db	1682 REFVPSRDEIEALIHROEMTSTVYCHGGGSCKITINSHTJAGEVEKILRGLAMEDSRM	1741
Qy	1741 FALPEYNGHVDKALIESRTVAVADVLAKEFKLAATSEVGLPMKFYFYLXCFLDTDNVPKDS	1800
Db	1742 FALPEYNGHVDKALIESRTVAVADVLAKEFKLAATSEVGLPMKFYFYLXCFLDTDNVPKDS	1801
Qy	1801 VEFAPMFQOAHAVIHGHHPABEENLOVLAALRLLOYLOGDYTLAAVILPLEBEVYSIQRLK	1860
Db	1802 VEFAPMFQOAHAVIHGHHPABEENLOVLAALRLLOYLOGDYTLAAVILPLEBEVYSIQRLK	1861
Qy	1861 ARISOSTKTFPCERLEBKRTSFLTEGTLRSRFRTSVVRQKYEBOEMDMWIKEEVSSAR	1920
Db	1862 ARISOSTKTFPCERLEBKRTSFLTEGTLRSRFRTSVVRQKYEBOEMDMWIKEEVSSAR	1921
Qy	1921 ASIIDKWRKFOGMOEOMAKYMALIKEMPGYSTLPDVECKEGSPQELMVLGVSADAVS	1980
Db	1922 ASIIDKWRKFOGMOEOMAKYMALIKEMPGYSTLPDVECKEGSPQELMVLGVSADAVS	1981
Qy	1981 VYKRGSGRPLEVFOYEHILSFGAPLANTYKIVDERELLFETSEVYDVAKLMKAYISMLV	2040
Db	1982 VYKRGSGRPLEVFOYEHILSFGAPLANTYKIVDERELLFETSEVYDVAKLMKAYISMLV	2041
Qy	2041 KKRYSSTRSASSQSSSR 2057	
Db	2042 KKRYSSTRSASSQSSSR 2058	

```
RESULT 3
US-09-815-379-12
; Sequence 12, Application us/09815379
; Publication No. US20030073613A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: GERRITSEN, MARY
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 10716/35
; CURRENT APPLICATION NUMBER: US/09/815,379
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/191,134
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-379-12

Query Match      95.5%; Score 10247; DB 9; Length 2048;
Best Local Similarity 99.8%; Pred. No. 0;
Matched 1961; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FCLOGTRWVLRNGQHFPSTVNSCAGIYVFRDYGQVFTYKOSTITTHOKTAMPTNEE 60
DB 1 FCLOGTRWVLRNGQHFPSTVNSCAGIYVFRDYGQVFTYKOSTITTHOKTAMPTNEE 60
QY 61 GVDWMASTLHSGSIMYNLFQRYKKNQIMTYIGSILASVNPYQPIAGLYEPATMEQYR 120
DB 61 GVDWMASTLHSGSIMYNLFQRYKKNQIMTYIGSILASVNPYQPIAGLYEPATMEQYR 120
QY 121 RHUGELPHIFATANCYRCIMKRNOCILIKGEGAGTSTKILKPLSVISQOSLE 180
DB 121 RHUGELPHIFATANCYRCIMKRNOCILIKGEGAGTSTKILKPLSVISQOSLE 180
QY 121 RHUGELPHIFATANCYRCIMKRNOCILIKGEGAGTSTKILKPLSVISQOSLE 180
DB 121 RHUGELPHIFATANCYRCIMKRNOCILIKGEGAGTSTKILKPLSVISQOSLE 180
QY 181 LSLKEKTSCEVERALLESPPIMEAFNAKTVYNNSSRFQFVOLNICOKNIOGGRIVDC 240
DB 181 LSLKEKTSCEVERALLESPPIMEAFNAKTVYNNSSRFQFVOLNICOKNIOGGRIVDC 240
QY 181 LSLKEKTSCEVERALLESPPIMEAFNAKTVYNNSSRFQFVOLNICOKNIOGGRIVDC 240
DB 181 LSLKEKTSCEVERALLESPPIMEAFNAKTVYNNSSRFQFVOLNICOKNIOGGRIVDC 240
QY 241 ILSQNRVVRONGERNYHIFVALLAGLEHEEREEFYLSPTENYHYLNOSGVEDTISD 300
DB 241 ILSQNRVVRONGERNYHIFVALLAGLEHEEREEFYLSPTENYHYLNOSGVEDTISD 300
QY 301 QSSFRVITAMDMQSKSEVREVSRLAGILHIGITEFTTGAQAVSFTALGRSAELL 360
DB 301 QSSFRVITAMDMQSKSEVREVSRLAGILHIGITEFTTGAQAVSFTALGRSAELL 360
QY 301 QSSFRVITAMDMQSKSEVREVSRLAGILHIGITEFTTGAQAVSFTALGRSAELL 360
DB 301 QSSFRVITAMDMQSKSEVREVSRLAGILHIGITEFTTGAQAVSFTALGRSAELL 360
QY 361 GLDPTQLTALTORSMFLRGEELITPLNYQOAVDSRDSLAMALYACCFEVIKINSRIK 420
DB 361 GLDPTQLTALTORSMFLRGEELITPLNYQOAVDSRDSLAMALYACCFEVIKINSRIK 420
QY 361 GLDPTQLTALTORSMFLRGEELITPLNYQOAVDSRDSLAMALYACCFEVIKINSRIK 420
DB 361 GLDPTQLTALTORSMFLRGEELITPLNYQOAVDSRDSLAMALYACCFEVIKINSRIK 420
QY 421 GNEDEFSIGLIDFGFENFVNHFEOFNINYANEKIQEFNKGITFLEOLEYSRBEGLVWE 480
DB 421 GNEDEFSIGLIDFGFENFVNHFEOFNINYANEKIQEFNKGITFLEOLEYSRBEGLVWE 480
QY 421 GNEDEFSIGLIDFGFENFVNHFEOFNINYANEKIQEFNKGITFLEOLEYSRBEGLVWE 480
DB 421 GNEDEFSIGLIDFGFENFVNHFEOFNINYANEKIQEFNKGITFLEOLEYSRBEGLVWE 480
QY 481 DIDMDINGECGLDIEKKGLGALLINEESHFPOATDSTLEKLSQHANNHFYKPPVAAN 540
DB 481 DIDMDINGECGLDIEKKGLGALLINEESHFPOATDSTLEKLSQHANNHFYKPPVAAN 540
QY 541 NFGVKAHAGEVOQDVAGILEKNDTFRDULNLRESRFDPIYDLFEHYSSRNNOQTLKC 600
DB 541 NFGVKAHAGEVOQDVAGILEKNDTFRDULNLRESRFDPIYDLFEHYSSRNNOQTLKC 600
QY 541 NFGVKAHAGEVOQDVAGILEKNDTFRDULNLRESRFDPIYDLFEHYSSRNNOQTLKC 600
DB 541 NFGVKAHAGEVOQDVAGILEKNDTFRDULNLRESRFDPIYDLFEHYSSRNNOQTLKC 600
QY 601 GSGHRRPTVSSQFVDSLHSLMATTSSNPFVRCIKPMNQKMPDQFDOAVVLNQARYSG 660
DB 601 GSGHRRPTVSSQFVDSLHSLMATTSSNPFVRCIKPMNQKMPDQFDOAVVLNQARYSG 660
QY 661 METVYIRKAGYVRRPQDFYKRYKVLBNLALPEDVQKCTSLLOLYDASNSEWOLK 720
DB 661 METVYIRKAGYVRRPQDFYKRYKVLBNLALPEDVQKCTSLLOLYDASNSEWOLK 720
QY 661 METVYIRKAGYVRRPQDFYKRYKVLBNLALPEDVQKCTSLLOLYDASNSEWOLK 720
DB 661 METVYIRKAGYVRRPQDFYKRYKVLBNLALPEDVQKCTSLLOLYDASNSEWOLK 720
```

```
QY 721 TKVFLRESLEQKLEKRRREEVSHAAVTRAHVGLFARQYKRVLYCVIITOKNYRAFL 780
DB 721 TKVFLRESLEQKLEKRRREEVSHAAVTRAHVGLFARQYKRVLYCVIITOKNYRAFL 780
QY 781 RRRFLHLKKAIVFOKOLRGQIARVRYROLAKREKOEKKOEKKOEKKOEKKOEKKOE 840
DB 781 RRRFLHLKKAIVFOKOLRGQIARVRYROLAKREKOEKKOEKKOEKKOEKKOEKKOE 840
QY 841 ERREAELRAQEEERKQOELEALOKSQKEALITRELEKOKENKQVEETILREKEIEDQ 900
DB 841 ERREAELRAQEEERKQOELEALOKSQKEALITRELEKOKENKQVEETILREKEIEDQ 900
QY 901 RKKEQOELSITBASLOKQERRDOELRLREBACRAAOEFLESINPDEIDECVNIERSL 960
DB 901 RKKEQOELSITBASLOKQERRDOELRLREBACRAAOEFLESINPDEIDECVNIERSL 960
QY 961 SGGSEFSSFLASACEKRNPNFSQPYREBEVDEGFADDDAFKQSPNPSHGHDDQRTS 1020
DB 961 SGGSEFSSFLASACEKRNPNFSQPYREBEVDEGFADDDAFKQSPNPSHGHDDQRTS 1020
QY 1021 GTRTDDSEEDPYMNDTVPTSPADSTVLLAPSVODSGSLHNSSGESTYCMFQONAGD 1080
DB 1021 GTRTDDSEEDPYMNDTVPTSPADSTVLLAPSVODSGSLHNSSGESTYCMFQONAGD 1080
QY 1081 LPSPDGDYDODDYEDGALITSGSSVTSNSYGSQMSPDYRCSVGTYNSSGAYRFSSEGA 1140
DB 1081 LPSPDGDYDODDYEDGALITSGSSVTSNSYGSQMSPDYRCSVGTYNSSGAYRFSSEGA 1140
QY 1141 QSSFEDSEEDPDSRDTDELSTYRDSYSCYTLTFYHFSFLYMKGLNMSMRRCVCLKD 1200
DB 1141 QSSFEDSEEDPDSRDTDELSTYRDSYSCYTLTFYHFSFLYMKGLNMSMRRCVCLKD 1200
QY 1201 EFTFLPFRSQOELKQGMHLKKGGSSTLSRRMKKRWFLRQSKLMPFENDSEELKGTV 1260
DB 1201 EFTFLPFRSQOELKQGMHLKKGGSSTLSRRMKKRWFLRQSKLMPFENDSEELKGTV 1260
QY 1261 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
DB 1261 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
QY 1321 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
DB 1321 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
QY 1321 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
DB 1321 EYRTAKEIINDNTKENGIDIIWADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHD 1320
QY 1381 SKGDRTRVEGEFTIVRGMLHKEVYKNSPKMSLSLKKRWFLVLTNNSLDYYSSEKNAKLCT 1440
DB 1381 SKGDRTRVEGEFTIVRGMLHKEVYKNSPKMSLSLKKRWFLVLTNNSLDYYSSEKNAKLCT 1440
QY 1441 LVNLISLCSVVPDEKIPKETGYMNTVYGRKHCVRLYTTGLNEATRMSVIONVDTKAP 1500
DB 1441 LVNLISLCSVVPDEKIPKETGYMNTVYGRKHCVRLYTTGLNEATRMSVIONVDTKAP 1500
QY 1501 IDTPPOQLIODIKENCLNSDVVEQIYKRNPIILRYTHHPHPSLPLPYGIDINLNLKDXG 1560
DB 1501 IDTPPOQLIODIKENCLNSDVVEQIYKRNPIILRYTHHPHPSLPLPYGIDINLNLKDXG 1560
QY 1561 YTTLODEAIKIPNSLOQLESMSDPIITIGILQTHDRLPLDELYCQLIKQTNKVPHG 1620
DB 1561 YTTLODEAIKIPNSLOQLESMSDPIITIGILQTHDRLPLDELYCQLIKQTNKVPHG 1620
QY 1621 SVGNLYSQWILTLCLSTFLPSRGIILKYLKFLHKLRIEOPGTMEKXALFTYESLKTVC 1680
DB 1621 SVGNLYSQWILTLCLSTFLPSRGIILKYLKFLHKLRIEOPGTMEKXALFTYESLKTVC 1680
QY 1681 REFVPSRDEIEALHROEYTSIVYCHGGSCKITINSHTTAGEVVEKLIRGLAMEDSRM 1740
DB 1681 REFVPSRDEIEALHROEYTSIVYCHGGSCKITINSHTTAGEVVEKLIRGLAMEDSRM 1740
QY 1741 PALFENGHVDAIRSRIVAVDLAKPEKLAATSEVQPMKFFLYCYFLDPTDNPVKDS 1800
DB 1741 PALFENGHVDAIRSRIVAVDLAKPEKLAATSEVQPMKFFLYCYFLDPTDNPVKDS 1800
```



QY 1801 VEFAPMEQAEBAVIGHHHPABEENLOVALAARLOQLQGYTLTHAIPLEEVYSIORLK 1860  
| | | | |  
DB 1801 VEFAPMEQAEBAVIGHHHPABEENLOVALAARLOQLQGYTLTHAIPLEEVYSIORLK 1860  
QY 1861 ARISSSTKTFPCERLEKRTSFLEGTLRSPRTGSVVRCKYBEOQMLDMWIKKEYSSAR 1920  
| | | | |  
DB 1861 ARISSSTKTFPCERLEKRTSFLEGTLRSPRTGSVVRCKYBEOQMLDMWIKKEYSSAR 1920  
QY 1921 ASIIRKMRKFGQMOEQAAMAKYMALIKEMPGYSTLFDEYCKEG 1964  
| | | | |  
DB 1921 ASIIRKMRKFGQMOEQAAMAKYMALIKEMPGYSTLFDEYCKEG 1964  
RESULT 4  
US-09-764-868-647  
; Sequence 647, Application US/09764868  
; Patent No. US2002016871A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 647  
; LENGTH: 1031  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (636)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-647  
Query Match 50.2%; Score 5385; DB 9; Length 1031;  
Best Local Similarity 99.5%; Pred. No. 1,2e-303;  
Matches 1021; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1032 DRYNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLSPDDGYD 1091  
| | | | |  
DB 6 DAWNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLSPDDGYD 65  
QY 1092 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 1151  
| | | | |  
DB 66 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 125  
QY 1152 DSRPDTDELRYRDSVYSCVTLPYFHSFLYMKGGIANSWKRWCVLKDETFLEWFSKOE 1211  
| | | | |  
DB 126 DSRPDTDELRYRDSVYSCVTLPYFHSFLYMKGGIANSWKRWCVLKDETFLEWFSKOE 185  
QY 1212 ALKQGMHLKGGGSSSTLSRRNWKGMFVLROSKLMEFENDSEBKLKGYEVRTAKEIIDN 1271  
| | | | |  
DB 186 ALKQGMHLKGGGSSSTLSRRNWKGMFVLROSKLMEFENDSEBKLKGYEVRTAKEIIDN 245  
QY 1272 TTKENGIDILIMDRFTFLAESPEDASQFVLSQVHASTDDEIOEMHDEQANPQAVGT 1331  
| | | | |  
DB 246 TTKENGIDILIMDRFTFLAESPEDASQFVLSQVHASTDDEIOEMHDEQANPQAVGT 305  
QY 1332 LDVGLIDVSCADSPDPSPNSFVITANRVLHCNADTPEEMHMITLLQSKSDTRVEGQE 1391  
| | | | |  
DB 306 LDVGLIDVSCADSPDPSPNSFVITANRVLHCNADTPEEMHMITLLQSKSDTRVEGQE 365  
QY 1392 FIVRGMLHKEVKNSPKMSLKLKGMFVLTHNSLDYKSSSEKALKGLVINSLSQVVP 1451  
| | | | |  
DB 366 FIVRGMLHKEVKNSPKMSLKLKGMFVLTHNSLDYKSSSEKALKGLVINSLSQVVP 425  
QY 1452 PDEKIFKETGYNNVTVYGRKHCYRLTYTKLNEATRWSSVIONVTDTPKAPIDPTQOOLID 1511  
| | | | |  
DB 426 PDEKIFKETGYNNVTVYGRKHCYRLTYTKLNEATRWSSVIONVTDTPKAPIDPTQOOLID 485  
QY 1512 IKENCLNSDVEBOIYKRNPIRLRYTHPLHSPLLPLPYGDIINLNLKDKGYTLLODEAIKI 1571  
| | | | |

DB 486 IKENCLNSDVEBOIYKRNPIRLRYTHPLHSPLLPLPYGDIINLNLKDKGYTLLODEAIKI 545  
| | | | |  
QY 1572 FNSIQOUESMDPPIIOGIIQOTGHDRLRLADELYCOLIKOTNVYPHFGSGNLYSMQIL 1631  
| | | | |  
DB 546 FNSIQOUESMDPPIIOGIIQOTGHDRLRLADELYCOLIKOTNVYPHFGSGNLYSMQIL 605  
| | | | |  
QY 1632 TGLSCTFLPSRGLIKYLLKFLKRIREOPPTGEMEKYALFTYESLJKTCKRSPVSRDIE 1691  
| | | | |  
DB 606 TGLSCTFLPSRGLIKYLLKFLKRIREOPPTGEMEKYALFTYESLJKTCKRSPVSRDIE 665  
| | | | |  
QY 1692 ALIHRQEMSTVYCHGGGCKITINSHTTAGEVVEKILRGLAMEDSENMPALPEYNGHD 1751  
| | | | |  
DB 666 ALIHRQEMSTVYCHGGGCKITINSHTTAGEVVEKILRGLAMEDSENMPALPEYNGHD 725  
| | | | |  
QY 1752 KAISRTVAVDLAKFEKELATSEVGDLPKMFYPLVYCFDLDNVPKOSVPAFMEQAH 1811  
| | | | |  
DB 726 KAISRTVAVDLAKFEKELATSEVGDLPKMFYPLVYCFDLDNVPKOSVPAFMEQAH 785  
| | | | |  
QY 1812 EAVIGHHPABEENLOVALAARLOQLQGYTLTHAIPLEEVYSIORLKARISOSTKFT 1871  
| | | | |  
DB 786 EAVIGHHPABEENLOVALAARLOQLQGYTLTHAIPLEEVYSIORLKARISOSTKFT 845  
| | | | |  
QY 1872 PCRLEKRTSFLEGTLRSPRTGSVVRCKYBEOQMLDMWIKKEYSSARASIIDKMRKQ 1931  
| | | | |  
DB 846 PCRLEKRTSFLEGTLRSPRTGSVVRCKYBEOQMLDMWIKKEYSSARASIIDKMRKQ 905  
| | | | |  
QY 1932 GMDQEAAMAKYMALIKEMPGYSTLFDEYCKEGFPOLMIGVADAVSVYKRGGRLE 1991  
| | | | |  
DB 906 GMDQEAAMAKYMALIKEMPGYSTLFDEYCKEGFPOLMIGVADAVSVYKRGGRLE 965  
| | | | |  
QY 1992 VFQYEHILSPGAPLANTYKIYVDERELFETSEVVDVAKLAKAYISHIVKRYSTTRBAS 2051  
| | | | |  
DB 966 VFQYEHILSPGAPLANTYKIYVDERELFETSEVVDVAKLAKAYISHIVKRYSTTRBAS 1025  
| | | | |  
QY 2052 SQGSSR 2057  
| | | | |  
DB 1026 SQGSSR 1031  
| | | | |  
RESULT 5  
US-09-764-868-1073  
; Sequence 1073, Application US/09764868  
; Patent No. US2002016871A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1073  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-868-1073  
Query Match 16.7%; Score 1789; DB 9; Length 348;  
Best Local Similarity 99.1%; Pred. No. 5.6e-96;  
Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1032 DRYNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLSPDDGYD 1091  
| | | | |  
DB 6 DAWNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLSPDDGYD 65  
QY 1092 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 1151  
| | | | |  
DB 66 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 125  
| | | | |  
QY 1152 DSRPDTDELRYRDSVYSCVTLPYFHSFLYMKGGIANSWKRWCVLKDETFLEWFSKOE 1211  
| | | | |

Db 126 DSRFDTDELVSRYRDSVSCVTLPRFHSFLYMKGGLMNSWKRRWCYLKDETFLMFRSKOE 185  
 QY 1212 ALKQGLHKKGGSSSTLSRRNMKKRWFLVROSLKMTFENDSEKLGTVETRAKELIIN 1271  
 Db 186 ALKQGLHKKGGSSSTLSRRNMKKRWFLVROSLKMTFENDSEKLGTVETRAKELIIN 245  
 QY 1272 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLSQVHASTDQEIQEMHDEQANPQNAVGT 1331  
 Db 246 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLSQVHASTDQEIQEMHDEQANPQNAVGT 305  
 QY 1332 LDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPE 1369  
 Db 306 LDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPE 343

RESULT 6  
 US-09-955-999-101  
 ; Sequence 101, Application US/09955999  
 ; Publication No. US20030036505A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barash et al.  
 ; TITLE OF INVENTION: Signal Transduction Pathway Component Polynucleotides, Polypeptide  
 ; TITLE OF INVENTION: Antibodies, and Methods Based Thereon  
 ; FILE REFERENCE: F1066P1  
 ; CURRENT APPLICATION NUMBER: US/09/955,999  
 ; PRIOR FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: 60/234,997  
 ; PRIOR FILING DATE: 2000-09-25  
 ; NUMBER OF SEQ ID NOS: 139  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 101  
 ; LENGTH: 348  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-955-999-101

Query Match 16.7%; Score 1789; DB 9; Length 348;  
 Best Local Similarity 99.1%; Pred. No. 5,6e-96;  
 Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1032 DRYMNTTVPTSPASSTVILLASVQDSGLHNSGSGESTYCMPOAGDLPSFDGDYD 1091  
 Db 6 DAWVNDTVPTSPASSTVILLASVQDSGLHNSGSGESTYCMPOAGDLPSFDGDYD 65  
 QY 1092 QDDYEDGATISGSGSVTFNSYSGQSPDYRCVGTYNSSGAYRFSSEGAQSSPEDSEDP 1151  
 Db 66 QDDYEDGATISGSGSVTFNSYSGQSPDYRCVGTYNSSGAYRFSSEGAQSSPEDSEDP 125  
 QY 1152 DSRFDTDELVSRYRDSVSCVTLPRFHSFLYMKGGLMNSWKRRWCYLKDETFLMFRSKOE 1211  
 Db 126 DSRFDTDELVSRYRDSVSCVTLPRFHSFLYMKGGLMNSWKRRWCYLKDETFLMFRSKOE 185  
 QY 1212 ALKQGLHKKGGSSSTLSRRNMKKRWFLVROSLKMTFENDSEKLGTVETRAKELIIN 1271  
 Db 186 ALKQGLHKKGGSSSTLSRRNMKKRWFLVROSLKMTFENDSEKLGTVETRAKELIIN 245  
 QY 1272 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLSQVHASTDQEIQEMHDEQANPQNAVGT 1331  
 Db 246 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLSQVHASTDQEIQEMHDEQANPQNAVGT 305  
 QY 1332 LDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPE 1369  
 Db 306 LDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPE 343

RESULT 7  
 US-09-851-682A-1  
 ; Sequence 1, Application US/09851682A  
 ; Patent No. US20020091248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Adams, Arwen E.  
 ; APPLICANT: Chiu, Choi Yung  
 ; APPLICANT: Duhli, David

; APPLICANT: Gorman, Susan W.  
 ; APPLICANT: Leng, Song  
 ; APPLICANT: Sheffield, Val  
 ; APPLICANT: Welch, Juliet  
 ; TITLE OF INVENTION: MYOSIN IYA AND CYCLIC NUCLEOTIDE GATED  
 ; TITLE OF INVENTION: CHANNEL-15 (CNCC-15) POLYNUCLEOTIDES, POLYPEPTIDES,  
 ; FILE REFERENCE: 200130.442  
 ; CURRENT APPLICATION NUMBER: US/09/851,682A  
 ; PRIOR FILING DATE: 2001-05-08  
 ; PRIOR APPLICATION NUMBER: US/09/172,422  
 ; PRIOR FILING DATE: 1998-10-14  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2548  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 US-09-851-682A-1

Query Match 13.9%; Score 1492; DB 10; Length 2548;  
 Best Local Similarity 24.3%; Pred. No. 1.2e-77;  
 Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;

QY 35 YGQVETTKOSTITTHOKVTA--MPTNBEQVDNMASTLHGGSIMYNLFQRYKNOIWT 91  
 Db 118 YGSLQSWLRVTERRRMRMERGFLPQPOQKDFDCLSPDLNEKTLLENLRDRFKKEXIYT 177  
 QY 92 YIGSLASVNPQPIAGLYEPATMEQYRRHGLPHIIPAIANECYCLMKRHNOCIL 151  
 Db 178 YVGSILIIYINPEK-FLPIYNPYVMYDNHOLQKPEPHIYAVADVAYAMQKKNQCIV 236  
 QY 152 IKGSEAGTBEETKILKFLSVISQOSLELSUKETSCVERALLESSPIMEAFNAKTVY 211  
 Db 237 ISGESGSGKTQSTNLIHLTLISQKF-----ASGEQIILGAGVLEHFGAKTAH 289  
 QY 212 NNNSSRFQFVQNLICQKNIQCGRIVDICLSSQNRVVRQNGERNYHIFALLAGLEHE 271  
 Db 290 NNNSSRFQFQIVNQGRTGLGA-YVEKYLLEKRLVYQEHNEBNHVFYLLGAGED 348  
 QY 272 EREBEYLTPEYHYHLNO-----SGVEDEK---TISDOE---SPREVTANDV 313  
 Db 349 EBSAHLQPEBYHYHLNOITKKPLRQSWDYCYSEPDQFVTEGDLRHDEERLOLAMEM 408  
 QY 314 MOPSREBEVRSRLAGILHIGNI---EFTAGGAOVSPYALGSALELGLDPTOLT 369  
 Db 409 VGFLEKTRQIFSLSALILHIGNICYKKKTYRDSIDICNEBVLPIVEELLEVEKEMLE 468  
 QY 370 ALTORSMFLRGEELTPLVQOAVDSRDSLAMALYACCFEYVYKKINSRIKGNEDFK-- 426  
 Db 469 ALVTRKTYVGEKILPYKLABAVTRNSMAKSLYSALFDMIVFRINHALNSKOLEHNT 528  
 QY 427 ---SIGIIDIGFEFENFEVNHFEQFININANEKLQYFKNHIFSLQLEYSRGLWEDID 483  
 Db 529 KTLSTGIVDIFEFEDYENNSFEQFINANERLQHYFOQHFKLEQBEYRREGISWNIID 588  
 QY 484 WIDNEGCDLIEKK-LGLLALINESHPPOATDSTLLEKLSQHANFHYVPRVAVNPF 542  
 Db 588 YIDNICTNLISKRTGTLHLIDESNFPQATNOTLLKPKQKHEDNSYIERPAVMEPAF 648  
 QY 543 GVKHYAGEVOYDVGRILEKNRDTFRDDLNLNLRBSRFDPIYL----- 585  
 Db 649 IKHYAGVKYGVQDFREKNTDHRKPDIVALLRSKNFISGMIGIDPVAVFWALIRAF 708  
 QY 586 -----FEHVSSRN-----NODTLKCG----- 601  
 Db 709 FRAMVAFREAGKRNIHRTGHDTPACAILKMSDFSFLQHPVQORSLEIIOCKEKYXS 768  
 QY 602 ----- 601  
 Db 769 ITRKNPRTPLSDLOGMALNEKNOHDTFDIAMNGRTGIRQSRLLSGTSLDKDGIFFANST 828  
 QY 602 -----SKHRRP 607

```

Db      829 SKLLERAHGILLTRNKNFKSPALPRGILLVNSLKLTLTLTQDRITKSLHLHKKKKRP 888
Qy      608 TVSSQFKVDSLHSLMATTSSNPNFVRCIKPNQKMPDOQDOAVYVNLQRLSGMETVRI 667
      889 S1SAPQQA-SLKLMTLTLQOAPRYFVKCIRSNABKULPLRFSDVYLQRLRTGMLTVOI 947
Qy      668 RKAGYAVRRPFDQFKRYKYLMLRNALPBDVGRKCTSLCLVDASNSEWQGLKTVFLNE 727
      948 RQSGYSKYSFQDFVSHFVLLPRNIPIS--KFNIDPFKXINLNDVNOQVKTWVFLKE 1005
Qy      728 SLEQLKREKEEV-----SHPAVITRAHVTGFLARQYRK- 763
      1006 QERHLODLHQEVLRRILLQRPFRVLLCRQHFHLRQPSVILQRPWNYINQKQVRDA 1065
Qy      764 -----VLYCVVITQKRYAPFLRRRFLHKKAIIVCQKQRLQGLAR----- 805
      1066 AVQKAPWASAAALQASWRAHLEKQYLEPAAAVIQCQKRDYRRHRMAALICQAR 1125
Qy      806 --VYQOLAKEKQOEKKEKKEEERERERERERERERERERERERERERERERERER 857
      1126 WKAYNE--SKRYQEQRKIILLQSTCRGFRARQRFKALKEQRLRTRKEVGLVNIKGY 1182
Qy      858 QOELEALQSQ--KEAELTRELEKOKENKQVEILIEKAI EDLQRMKE---QOEISLTE 912
      1183 SLEIGSDPSEWEDSPDNRIRKAIIECKSVIESNRISRE--SSVDCIKESPNKQOERAQSQ 1241
Qy      913 ASLOKLOE-----RRDOELRLREBECRAOEFLESLNTEIDECRANE-----SLSGG 963
      1242 SCVD--LOEVDVLYREBRSLLEDLHQKVGRAKR--BSRMELELOALPSLELLKVNLSGSI 1298
Qy      964 S-----EFSSELAESACEE--KPNFNSQ-----PYDEER-----VDEGEFADDD 1001
      1299 SPSEDRMWTETELVPEGLQSPGTPDESQSLLESLYSQSKLSEVTSDEG-----D 1353
Qy      1002 AKQDSPNBEHGHSDORTSGIRTSDDSGE---EDPYANTVYVPPSPADSTVVLAPSVQ 1057
      1354 LQPPPKLSSSPKFSRDNALASNETSAAHLKQGTMKEMVVCSEES---ITCKPOLK 1409
Qy      1058 DSGSLHNSSGESTYCMQPNAGDLPSPDDYDYDODDYICAITSSSVTFPSNSYGSQMS 1117
      1410 DS---FISHSLPTFFIQQODPLKTNLSQJDTSIQRKLLINEDYALTLIDINETR-- 1464
Qy      1118 PDYRCS---VGTVNSGAYRFSSEGAQSFEDSEEDFSRDTDEL----- 1161
      1465 -RYHSGSKQOIVPSLNTBSSNPVLKLELNTKEKERQKQLQOQNKEMMEQIROQTDL 1523
Qy      1162 -----SYRDSYVSCVTLFYFHSFLYMKGLANSWKIRWCVLKDETFMLFRSKQE-- 1211
      1524 EKERRAFKTIERPRIGECLVAP-----SSYQSKORVERPSILSLNTSNKGE 1570
Qy      1212 -----ALKQGMHKGGGSSSTLSRNMKKMFVLROKILMYFENDSEKLGTV---- 1260
      1571 LNVLSLSLKDALALQKSSSAHLPPKD-----REVTVFPERKSGPCOSSVYKELS 1621
Qy      1261 -----EVRTAKEILDN-----TTKEN-GIDI 1280
      1622 KTDNMGTOLVNACKLSNNRISKREHRTPTOSYSHNSDDLREGNARPIFTKDNMSTIL 1681
Qy      1281 IMADRTFHILAESPP---EDASQW--FSVLSOVHASTDQHIOEMHDEQANPONAVGTLDV 1334
      1682 VEKEA---LNSKRPQLHKEDEBAMKPVKLADGQRETQSOFSSV--DEQAKLHK--TMSQ 1734
Qy      1335 GLIDSYC---ASDSDRPNSPVITIANVNLHGNADTPEHMHWTITLQSGDPRVE-- 1388
      1735 GEITTKLAVQKASDSQDIRPOR-----ALMFPWAKGQGEKKTTRVKT 1777
Qy      1389 -----GOEFI-----VRGMLKEVANSFMSLSKLKKRW 1417
      1778 TOSEVSPPLFAGTDVIRAHQFPBELAYHPTPLSPPLPSCKREFENKEPBP----- 1830
Qy      1418 PVLTHNSLDYKSSSENAKLGLTLVNSICSVVPPDEKIFKETGVNNTVTVGRKHCYRLY 1477

```

```

Db      1831 -----KAKKRSVKISNVALDSM-----HWQ----- 1851
Qy      1478 TKLINEATRMSSVIONVDTFRAPIDTPTQOLIODI-KENCINSVVEQIYKRN-PILRYT 1535
      1852 -----NDSVQ---ILASVDDLKS--HDEFLLKRVNLDNEDSKKDLVDVFPKALKEFRQN 1903
Qy      1536 HHPLHSPLLPLPYGDININLLKDKGYTTLQDEAIKIFNSIQOLBSMSD-P1P1IOGIIQT 1594
      1904 IFSFYSALAMDG---KSIRYKDLVALFEQILEKTRLEQRDSLCESSPVAV-----W 1953
Qy      1595 GHDLRPLADELYCOLIKQTN---KVP-----HRSVGNLYSQIILT- 1632
      1954 VNTFVKPLDE--YMEFKTSDCTATKVPTEERRKRRKETDVEHNGHAI PRATOYSIPTY 2012
Qy      1633 CLSCTFLPSRGLIKLKEFKLRIREOPGTEMEKVALPTYESLKT-KC-----R 1681
      2013 CEYCS-----SLWINDRAVCULCKYACHKKCLKTTACSKKYDELSR 2059
Qy      1682 EF-----VPSRDE-----IEALIHROEM---TSTVYCHGGGSCKITINSHTTAGEVVE 1726
      2060 QGVELSRLTSEDRTVPLVEKLIYIMHGLYTEGIVRKSGSTWKI-----K 2107
Qy      1727 KLIRGLAME-DSRNFALFEYNGHVDKAIESTVAVADVLAFBKLAATSVDLPMKRY- 1784
      2108 ELRQGLDIDDAESVN---LDDYNIH-----VIASVFKQW-----LRDLPNPLMT 2147
Qy      1785 FELY 1788
      2148 FELY 2151
Db

```

```

RESULT 8
US-10-177-293-315
; Sequence 315, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Kamatekar, Shubhangi
; APPLICANT: Mettens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 315
; LENGTH: 1855
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-177-293-315

Query Match      13.6%; Score 1461; DB 9; Length 1855;
Best Local Similarity 25.6%; Pred. No. 4.9e-76;
Matches 516; Conservative 350; Mismatches 681; Indels 468; Gaps 78;

QY 7 RWLARENGHFSTVNSCAE-----GIVFRIDYQGVFTYKOSTITHQVMTMHP1 57
DB 12 RWIIP-----PEEVMKSAELKDYKPGDKVLLHLEBKDELYHLDPKTELPHLRNDP 66
QY 58 NEBGVDMASTLTHGSGIMWNLFOYKRNQ-IMTYGSIASVNPYPIAGLYEPATME 116
DB 67 ILVGENDLRLSLHBAVLAHNRVRFIDSKLYTTGCVLVANINYEQLP-IYGEDITN 125
QY 117 QYSRRLGELPRHIFAIANECCYALMKRHDNOCILIKGSGAGCTESTKILKFLSVISQ 176
DB 126 AYSQNMGMWDPHIFAVAEAYKQMARDEBNOSIIVSGSGAGCTVSAKYAMKFPATVSG 185
QY 177 QSIHELKTEKTCVBEBAILESSPIAEAFNAKTVNNNSRBEKFPQOLNCOGNIOGGR 236
DB 186 SAEAAVEEK-----VLTSNPMESIGAKTRDNNSSRFQKYLEIGFDKRYRITIGAN 238
QY 237 IVDCILSSQNRVVRONPGERNYHIFYALLAGLEHEEREFEYLSSTPENHYLNOSGVEDK 296
DB 239 MKRYLL-EKSRVFOABEENYHIFYOLCASALPEFKRLGNADNPNYTKOGSSPVLE 297
QY 297 TISDQSFREVIAMDVMOFSKEEVBESKLLAGILHLGNIETAGAAVSF--KTAL 353
DB 298 GVDADKEMATROACTLLGISESHQMGIFRILAGILHLGVGFTSRDADSCITPRKHBP1 357
QY 354 GRSAELGLDPTQLTALTORSMFLGEBELTPLTNQOANDSDSLMALYACCEFWVK 413
DB 358 CIFCDLMDGVDEBCHMLCRKLATATEYIKPISKLQATNADALAKHIALYALFWIWD 417
QY 414 KINSRI-KGNEDEKSGILDIFGFENFEVNHFEQFINYANEKLOEYFNKHFISLEOLEY 472
DB 418 NNQNALHSAVKQHSFIVGDIYGFETFEINSFEOFCINANEKLOOQFMNHYKLEQBEY 477
QY 473 SREGVAMEDIDWIDNECDLIEKLLGLALLINESHPOATDSTLLEKLSHANN-HF 531
DB 478 MKEQIDPWLIDFYDNQPCILIESKLGILDLDEBCKMPKGTDTYMAQKLYNTHLKCAL 537
QY 532 YVYERAVANNFGVYHAYAGEVQYDVRGILEKNDPTFFDOLLNLBESRPFYIDLEFVSS 591
DB 538 FEKPRLSNKAFTIIOHPADKVEYQCEGFLEKXKQTVFEEQIKVLSKSKFKKLPELF----- 592
QY 592 RNNQDTLKCS-----KHRRPTYSSQPKVDSLHSLMA 623
DB 593 ---QDEBKASIPFSATSSGRTPLTRTPAKPTKGRBQMAKENGKTYGHQFR-NSLLHLNB 648
QY 624 TLSSNPFVYRCIKPMQKMPDOQDAVLNQLRYSGLMLETVAIRAGYAVRRPQDFYK 683
DB 649 TLNATPRHYRCIKPNDPKLPFTEDEKRAVOQLRACQVLETIRISAAGPSRMTYQEPFS 708
QY 684 RYVNLNRNLALPBDVAGCTSLQLYDASNEBWLQTKVFLBESLEQLEKREBEVSH 743
DB 709 RYRVLMKKQKDVLSDRQTCGNVLEKILLDKQYQFKTKIFFRAGOVAYLEKLRADKLKA 768
QY 744 AAMVIRAHVGLFARQYRKVLVCV-----ITOKYVYAF 778
DB 769 ACIRIQKTIRGMLLRKKYLRMKCAITMQRVYNGYQARCYAKFLARTKATITIIQKWRY 828
QY 779 LRRRLHLKKAALVFOKOLQIARVYRQLA-----EKQOE 819
DB 829 VVRARYKIRRAATVLOSRLGLARNRYKILRGKAVIIQKRVGMLARTHYKKSMA 888
QY 820 -----KKQOEKREKREBEERER-----ERREALRAQOE----- 853
DB 889 IIVLOCCEFRMAKRELKKLKIARSVERVYKLRIGEMENKIMQLOKRVDEONKDYCLAVE 948

```

```

QY 854 -----ETRR-QOELEALQKSOKEABLT--RELEKQENKQVEILLREKIEDL 899
DB 949 KLTNLEGVINSBETKLRSDLERLOLSEBAKYATGRVLSIQ-----BEIALRLDLE-- 1000
QY 900 QRMKEQOELSTLSEALQKQERRDOELRLREBAERAAOEFLESLN--FDEIDCVANI 956
DB 1001 ---QTRSEKKECIEEADRYKQETEOJVSNLKEBNTLKQE--KEALNHRIVQAKMTEITM 1056
QY 957 ERSLSGSEFSESELASACEEKNF-----NSQPYPEBEVEGFEADODAKOSPNSPE 1011
DB 1057 EKKL---VEETKQLELDLNDERLRYQNLNEFSR--LEERYD-LKEEHTLMVHVPK-- 1108
QY 1012 HGHSDQRTSGIRTSDDS-----SEEDPYNDVTVPSPSADSTVLAPSV----- 1056
DB 1109 -GH-KRDSHTSSNESEVIFSESLAEAMEDIISRREBPSEKVPPLDMSLFLKQKRYTEL 1165
QY 1057 -QDSGLNHSNGSESTYCPQNGDLPSPD--GDYVD---QDYDEGAITSSSVTF 1108
DB 1166 EOEKQVMODELDRKEEOVLRSAKEEERPOLRGALEYESLKRQLESSENKKLKNELNEL 1225
QY 1109 SMSYSGMSPDVRCVGTNNSGAYRFSSGQSSFFDESEBDFDSRFTDDELSYRDSV 1168
DB 1228 RYALSEKSAPEV-----TAPGAPAYRVLMEOULTS-----VSEELDYRKEEV 1266
QY 1169 YSCVTLLPYHPSFLYMKGIMNSW-----KRWCVLKDETFILW-----FRSKQALKQGL 1218
DB 1267 -----LILRSQVSKKAIOPKDKNTMTDSTILLEDVQKMKDGE-IAQAVI 1313
QY 1219 H-KKGGSGSTLSRRNW--KRWFLV-----QSKLMYFENDSEKUKGYVE 1261
DB 1314 GKETNRSSALDYHNELEDGELMVLVEGLKQANRLLESOLQSKSHENEA-EALRG--E 1370
QY 1262 VRTAKEIIDNTKENGIDIMADRTFHLIAS--PEDASQSFVLSOVHASTDOEIOEM 1318
DB 1371 IOSLKE-----EN-----NRQOQLAQULQLPPEARLEASLOHETRLTNENDLM 1416
QY 1319 HDEQANPON-----AVGTLDVGLIDVCAAS--DSPDRPNSFVIITANRV 1360
DB 1417 -EOLEKQDKYTRKLKQOLKVPFAKKIGELEVGQMNISPGQIDBPFR----- 1463
QY 1361 LHCNADPBEEMHMTTLQRSKGDTRVSGOEFYVGMHLKEVYNSPKNSSLKLRK-- 1416
DB 1464 ---VNIRKEKDFQGMLEYKKEDEQ-----KLAKN--LILELKEFGYAV 1502
QY 1417 -----WFLVTHNSLDVYKSEKXALKGLTVLNSLCSVVPPEKIFKEGYNNV 1465
DB 1503 NLIPGLPAILYPMCVRH--ADYLANDQK-VNSLITSTINSIKVLAKKGDDFETYSFMLS 1559
QY 1466 TVYGRKHCVRLYTK-----LLEATRWSSVIONVDTKAPIDPTQOOLIODIKENCIN-- 1518
DB 1560 NTCRFLHCLKQYSGEGEMKNTSR-----QV-----EHLCTLPND 1594
QY 1519 -----SDVEQIYKKNPILRYTHNPLHSPLEP--LPYGD-I-NLNLKDKGY-----TT 1563
DB 1595 LAEYQVLDLAIQIYQQ--LVRVLENILQPMIVSGMLEHETIQGVSGKPGFLKRTSS 1652
QY 1564 LODEAIKIFENS-LQOLEMSDIPRIIOGLIQGHRLPYRDELVGOOLIKQNTKVNHPGSV 1622
DB 1653 IADESTYTLDSILRLQNSFHS-----VWCQHMDP--ELIKQVVKQMFYIIGAITL 1701
QY 1623 GNLYSQMLITLCSFPLSPRGILKYLKRIKRPQFSTE-MEKYALFTYESL----- 1675
DB 1702 NNL-----LKRKMCGW--SKGM-QIRNVQSGLEMLNDKXNLMSGAKETLEPLLOAQL 1753
QY 1676 ---KTKYCREFEVPSRDEIBALIHROEMTSYVYCHGGSCKI--TINSHTTAGAVEYKL- 1728
DB 1754 LQVKKKTD-----DDEAALCS-----MCNALTQAQIVKVLNLTYPVNEFEERVS 1797
QY 1729 ---IRGLAME-----DSRNFAL--FEYN 1747
DB 1798 VSFIRTIQRLRLDRKDSPOLMDAGHIFPVIFPN 1832

```

```

RESULT 9
US-10-205-823-273
; Sequence 273, Application US/10205823
; Publication No. US20030108963A1
GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamackar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatz, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Duolin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: PaateSeq for Windows Version 4.0
; SEQ ID NO 273
; LENGTH: 1285
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-205-823-273

Query Match      13.0%; Score 1391; DB 9; Length 1285;
Best Local Similarity 31.1%; Pred. No. 3.4e-72;
Matches 368; Conservative 225; Mismatches 365; Indels 226; Gaps 33;

QY 10 LBNGHFPSTVNSCAEIVFRITDYGQVFTYQSTITTKQKTAHMTPEBEGDMDAST 69
DB 34 LKQKGTPTALIN-----QVPAEED-----SKQVEDNCSTLM 66
QY 70 ELHGSIMYNLPOQRYKRNQIWTYIGSILASVNPYQPIAGLYEPATWEOYSRRHLGELPH 129
DB 67 YLNEATLNLINIKRYSKDRITTYVANILAVNYPPIPKIYSEAKISYQSGSLGRPH 126
QY 130 ITAIAINECYRLKWRHNDQCILIKGESGAGKTESTYLILKFLSVISQGSLESLKERTSC 189
DB 127 VFALADKARFDMKVLKMSQSIIVSGSGAGKTEHTKFLRYLTESYGTGD----- 177
QY 190 VERALLESSPIMEAFGNATKYVNNSSRGKPFQOLNIQKGNIGGRYIDCLISSONRV 249
DB 178 IDRIVEANPFLLEAFGNATKYVNNSSRGKPFVEIHFNKSSVVG- FVSHYLLKSRIC 236
QY 250 RONPGRNTHIFAYALLAGLEHEBEREFYSTPENYVYN----- 288
DB 237 VQGEKERNHIFRILCAGSSEDIREFKHLSSPDPNPFYLRGCTRYRANKETDKOILQNK 296
QY 289 -----OSGVEDKTISSQSFREVTAMDVMOFSKEEVEBSLLAGILHGNIEF---- 339
DB 297 SPEYLAGSGMKDPLDHDGDFIRMTAKKIGLDDEKLDLFRVAVGLHLGNIDPEBAG 356
QY 340 ITAGAGQVSPKTA--LGRSALLGLPTQLTALTLQFSMF-----LRGEIILPLNVQA 392
DB 357 STSGGCLNKKKSQSLSEYCAELLGLDQDRLVSLTTRVNLTTAGTGKTVIKVPLKVEQA 416

```

```

QY 393 VDSRDSLAMALYACCFFEWIKINSRIGKNEDPKSIIGLIDIFGFENFEVNHFEQFNINYA 452
DB 417 NNARDALAKTVYSHLFDHVVNRVNCQFPFETSYRIGVLIDLAGFEYFEHNSPEQCINVC 476
QY 453 NEKOEYFNKHLFSLQELYSREGIWMEDIDMIDGECGLDILIEKL-GILALINESHP 511
DB 477 NEKLOQFNERILKEQELYOKEGIVNHHVNDQCIDIBALTVGLIDILDEBNLP 536
QY 512 QATDSTLEKLSOHANHFYV-----KPRVAVN-----NFGVGYAGEVOYDVRGILE 560
DB 537 QPSDQHFTSAVHQK-KHFRLLTRPKSKLAVHRIKDDEGFIIRHFGANCYETETQVE 595
QY 551 KNRDTRFDLLNLARESRFDPIYDLFEVHSSRNODTLKCSKRRPPTVSQFYVDSLHS 620
DB 596 KNNDALHMSLESJLIESDKFIRELFES-STNNKNDOTQKXGKLSFISVGNKFKT-QNL 653
QY 621 LMATSSNPPFVRCIKRNQMOPQOAVVNLQRLRSCHLETYRIRKAGYAVARPPQD 680
DB 654 LLDKLRSTGASFIRCIKPNLKWTSHHFEGAQILSQCSGNVSVYDLMOGGYPSRASFHE 713
QY 681 FYKRYKVLNRNLALPEVYRGKTSLLQYDASNSMOLGKTVPFLRESLEOKLBRGRREE 740
DB 714 LYNNYKTKMPDLALDRLFCALFKALGLANENDYKGLTKVFRPGKPAEPDQIMKSD 773
QY 741 VSHAAMVIRAHVGLFLARKQYKVLVC-----VYIIQKYRAFLRR 783
DB 774 PDHLAEIVK-RVNHMLTCSRWKQVOWGSLSVYKLKNKIKYBAEACIKQKTRIMVLCRR 832
QY 784 FLH-----LKAALVF-----OKOLRG-----QIARRYROL 811
DB 833 --HKPRIDGLVKVGLTKRLDKFNEVSVLKDQKPENNKQIKNLEISIDTLMAKIKSTM 890
QY 812 AEKREOE-----KKQOE--EKKQREBERERERRRRAE--LRAQ 851
DB 891 TQEQIQEYDALVSSSEBLSALQKKQOEBAELRLRIQEBMESGRKREDEGRKRE 950
QY 852 BEETRKOQELALQSKQEAELITRELEKQKKNQVEEILREKEITDLQMKKEQOELS 911
DB 951 BEERMKLEMAKRRQSEBERKREDEKRIQAEVLAQLAKQE-----BESQQA 1003
QY 912 EASLOKQERRDQELRLREBEACRAAQEFLSLNDFEIDECVRIERSLSG----- 963
DB 1004 E-----OERRRREY-----ALRIQSEAEILIS---DEAQADALRRNOSTRPMPE 1047
QY 964 -----SPPSSR-----LAESACEEKNPFNFPQPYEEVEVDGFEADD----- 1001
DB 1048 QMAKEMSEPLSGPVLTKAAAGTKYDLSK-WKYAEALRDTINTSCDIBLLACREBPH 1106
QY 1102 -----AFKDSNPSEHGHSDORTSGIRTSDDSEEDPYMNDT 1038
DB 1107 RLKXYHAWK-SKNNKNTTETGQAPKSVTDYDFA---PFLNNS 1146

RESULT 10
US-10-177-293-317
; Sequence 317, Application US/10177293
; Publication No. US200301024128A1
GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatz, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamackar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzatzi, Lajos

```

APPLICANT: Meric, Funda  
APPLICANT: Sahin, Aysegul  
APPLICANT: Mill, Gordon B.  
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-038  
CURRENT APPLICATION NUMBER: US/10/177,293  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/299,887  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 60/301,572  
PRIOR FILING DATE: 2001-06-27  
PRIOR APPLICATION NUMBER: US 60/306,501  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US 60/325,002  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: US 60/362,585  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
PRIOR FILING DATE: 2002-05-14  
NUMBER OF SEQ ID NOS: 506  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 317  
LENGTH: 1285  
TYPE: PR  
ORGANISM: Homo sapiens  
US-10-177-293-317

Query Match 13.0%; Score 1391; DB 9; Length 1285;  
Best Local Similarity 31.1%; Pred. No. 3,4e-72;  
Matches 368; Conservative 225; Mismatches 365; Indels 226; Gaps 33;

10 LRENGHPESTVSCAGIIVFRTDYGQVFTYKOSTITQKVTAMHPTNEEGVDNASLT 69  
34 LNKGTFTLALIN-----QVPAEED-----SKDVEDSCSLM 66  
70 ELHGSIWYNLFGQRYKRNQIWTYIGISLVNFPYQIAGLYEPATMEQYSRRHGLRPH 129  
67 YLNEATLNLHNIKRYSGDRIYTVANILAVNPYPIPKYSEALIKSYQKSLGTRRPH 126  
130 IFPAINECTRLMKRDNOCILIKGSSGACKTESTLILKFLSVISQOSLESLKERTSC 189  
127 VFPAIAKAPFDMVLKMSOSIIVSGSGAGKTENTFVLRLYTESYGTQD----- 177  
190 VERALIESPIMEAFNGAKTVVNNSSRFQKVFQOLNCKGNIOGRIYDCILSSONRVV 249  
178 IDRIYEAFLPLAFNGAKTVRNNSRRGKFEIHFNEKSSVYGG-FPSHTLEKSRIC 236  
250 RQNGERNYHIFYALLAGLEHEEREEFYLTSPENHYLN----- 288  
237 VQGEKERNYHIFYRLCAGASEDIRKHLSSPDNFRYLNRGCTRFANKETDKQILQNRK 296  
289 -----OSGCVEDTIDQSEFRVITAMDMQFSKEVNEVSLLAGLILHGNIEF----- 339  
297 SPEYLAGSKMDLDDHGDIFIMCTAMKKGIDDEKDLDFRVVAGVHLGNDIEEAG 356  
340 ITAGAGVSPKTA--LGRSAELLGLDPTQTLDTLJORSMP-----LRGEILPLVWVOA 392  
357 STSGCGNLKMKKQSLFECHELGLDODLRVSLTRWMLTTGKRGKVIKPLVKEON 416  
393 VDSRDSLAMALVACCFEWMYIKKINSRIKGNEDKSGIGLIDIFGENFEVNHFEQFINYA 452  
417 NNARDAALAKTVYSHLPDHVNVNRNQCPEFTSSYFGLVIDIAFEYFEHNSPFCINC 476  
453 NEKLQGYFNKHTISLEQLEYSRGLYMEDIDWIDNECIDLKCU--GLLALINESHP 511  
477 NEKLQGYFNKHTISLEQLEYSRGLYMEDIDWIDNECIDLKCU--GLLALINESHP 511  
512 QATDSTLLEKLSCHANNHFFV-----KPRVAVN-----NFGVYAGBVOYDVAGIIE 560  
537 QSDQDFTSAVHOKH-KDHFRLTIPKSKLAIVARNRIRDBEGFIRHFAAGVCEETTFQVE 595  
561 KNRDTRFDLLNLRESRDFIYDLFEHVSRRNODTLKCGSKRRPPTVSSQKVDLSLS 620

596 KNDALHMSLESLICESRDKFIRELFE--STNNNDOTQKAGKSLFISVGNKFT-QNLT 653  
621 LMATSSSNPPFVRIRIKNMQMPQPOQAVVNLORRSGMLFETVRIKAGAVARPPQD 680  
654 LLDKLRSTGASFIRICIKNPKTSHHFEQAOLISLOCSGWSVLDLMOGGPSPASFPHE 713  
681 FYKRYKVMRNALALDEVDYRGKTSILQYDASNSSEWQKTKVFLRESLEQLEKREBE 740  
714 LYNNYKMYKPDRLARLDRPLPCALFKALGNENDYKGLTKVFPKPKAFEPDQIMSD 773  
741 VSHAAMVIRAHVGLFARKQYRKVLVC-----VVITQKYRAFLLRR 783  
774 PDHLAELVK-RVNHMLTCSRMKQVQCSLSVYKLKXIKYPAEACIKMQKTIKRWLCRR 832  
784 FHH-----LKKRAIVF-----QKOLRG-----QIARVYRQL 811  
833 --HKPRIDGLVKGTLKRLDKFNEVSVLKDGRPNKQIKNLEISIDTLMAKIKSTM 890  
812 AKRQGE-----KKQGE--EKKRKEERERERERE--LRAQ 851  
891 TOEQIQEYDALVKSSEELLSALQKKQGEBAERLRIQEMEKERRREDEKRRKE 950  
852 EETRKQOEALQSQEABETRELEKOKENQOYEBILRLEKIEDIQMKQOELS 911  
951 EERERMKLEMEAKRQGEERERKREDEKRIQAEVLAQLOKE-----EESQOQAVL 1003  
912 EASLOKQERPOEALRLEBEACRAOEFLESINDEIDECVNIERSLSG----- 963  
1004 E-----BRRREL-----ALRTAQBELIS---DEQADALARNNDTRPKMTP 1047  
964 -----SEPSSE-----LAESACEEPNFPQOPYEEVDEGEFADD----- 1001  
1048 QWAKEMSEFLSGPAVLATKAAGTKYDLS-KWYALRLDTINSCIELLAACREBH 1106

1002 -----AFKOSPNSPEHSHSDQRTSGIRTSDDSEEDPYWNT 1038  
1107 RLKTVYHAWK-SKNKKRTETEQRAPKSVTDYDFA---PFLNNS 1146

RESULT 11  
US-10-044-303-2  
Sequence 2, Application US/10044303  
Patent No. US20020137161A1  
GENERAL INFORMATION:  
APPLICANT: Max-Planck-Gesellschaft e.V.  
TITLE OF INVENTION: Protein expression and structure solution using  
FILE REFERENCE: ST010209-BPA  
CURRENT APPLICATION NUMBER: US/10/044,303  
CURRENT FILING DATE: 2002-01-11  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1016  
TYPE: PR  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Whole sequence  
OTHER INFORMATION: The protein comprises as component (1) the aa  
OTHER INFORMATION: Sequence of myosin II motor domain of  
OTHER INFORMATION: Dictyostelium, a three aa linker region and the  
OTHER INFORMATION: a-actinin aa sequence  
US-10-044-303-2

Query Match 12.7%; Score 1360; DB 12; Length 1016;  
Best Local Similarity 34.3%; Pred. No. 1.6e-70;  
Matches 354; Conservative 207; Mismatches 360; Indels 112; Gaps 29;  
8 VLRENGHPESTVSCAGIIVFRTDYGQVFTY-----OSTITHQKVTAMHPTNEEGV 62  
39 IWYNDP-----PKERDYSYCGEIVSET--SDSFTFTVDGQDQVKKDDANQRNPIKFDGV 92

[illegible]

```

: Publication No. US20030087270A1
: GENERAL INFORMATION:
: APPLICANT: Schlegel, Robert
: APPLICANT: Chen, Yan
: APPLICANT: Zhao, Xumei
: APPLICANT: Monahan, John
: APPLICANT: Monahan, Shubhangi
: APPLICANT: Glatt, Karen
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Hoersth, Sebastian
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
: TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
: TITLE OF INVENTION: OF CERVICAL CANCER
: FILE REFERENCE: MFI-035
: CURRENT APPLICATION NUMBER: US/10/171,311
: CURRENT FILING DATE: 2002-06-12
: PRIOR APPLICATION NUMBER: US 60/298,159
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/298,155
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/335,936
: PRIOR FILING DATE: 2001-11-14
: NUMBER OF SEQ ID NOS: 238
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 164
: LENGTH: 1938
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-171-311-164

Query Match      12.4%; Score 1333; DB 9; Length 1938;
Best Local Similarity 24.2%; Pred. No. 1.3e-68;
Matches 514; Conservative 362; Mismatches 746; Indels 498; Gaps 72

      8  VMLRENGHF--PSTVNSCAGGIYFRTDYGQVFYTKOSTTHQKVTAMHPNREGVDM 65
      36  VWVSEKQGFMAAIKEKGBVVELVENG-----KVTYQKDDIQKMPFKSEKVBDM 90
      66  ASLTLEHGSIMYNLFOYKKNQIWTYIGSIILASVNPQPIAGLYEPATMEQSRHJGE 125
      91  AELTCINASTVLMHLRERYFSGLLYTTGSGLCVYNVPKHLR-IVSEKIVDMYKKGKHE 149
      126  LPHIIPAIANECYKCLMKRHNOCILKIGESGAGTSETKILKFLSVISQOSELSLKE 185
      150  MPPIYALADYAVYSMLQDRDOSILCTGSEGAGKTENTKKVIQYLAIVASSHKKQDTS 209
      186  KTSCEVERALLESSPIMEAFGNAKTYNNNSSRFGCVQNLNCOQKNIQGRIVDCILSSQ 245
      210  ITGEIEKQQLDANPLLEAFGNAKTYKDNSSRFGFIRINIDVTGIYIGANT-ETYLEK 268
      246  NRVVRONGENGRNIHYFALLAGLEHEEREFEYLLSTPENYHYLNOSGVEDKTIISQESFR 305
      269  SRAIRQARDERTFHFIFYMIGAKEKKRSDDLBSGFNNYTLIS-NGFVPIPAQDDMEQ 327
      306  EVITAMDVQFSKEEVRVSSLAGILHLGNIET--ITAGAGQVSPFALGRSAELLGL 362
      328  ETVBAMAIMGSEEBEQLSILKVSSVLDLGNIVFKKEKNTDQASMPDNTAAQKVGHLMI 387
      363  DPTQUT-DALTQSRMFLRGESEILIRPLNQAVD-SRBSLMAALVACCEFWYIKATNSR- 419
      388  NVTFETRSILPRIRKV-GRDVVQAQOTKEQADFAVEALAAKYERLRRWILTRNVKLD 445
      420  -KGNEDFSKISILIFGFENFEVNHFEQFNINVAEKKIOEFNKHIFISLEOLEYSGRELV 478
      446  KTHQGSFGLIDILDIAGFEIPEVNSFBDCLINTNEKIQQLPNNHTM-IFLDEQETREGIE 505
      479  WEDIDW-IDNGECLDLIEK--KLGLALINEESHFPQATDSTLEKLSHGHANNHFFVK 534
      506  WNFIDFGDLDPCTELIERPNNPGVALLDLDEECMFPRATKDSFVEKLTCEGSHPKFOK 565
      535  PRVAVN--NFEVQYIAGEVQYDVAGLEKQRDTRRDDLLNLBRSRPFPIIDLFEHVS-- 590
      566  PKQKDKTEFSEIHYAGVQVYNASAWMLTKNMDPLNDVYTSLLNASSDKFVADRLWDXDVR 625

```

```

Qy 591 -----SRNODTLKCGSKHRR---PTVSSQFKVDSLSHSLMATLSSSPFVACIKENMG 641
Db 626 VGLDQAKMTESSLPSSKTKKGMFRVGLYK-EQGLKMTLRLNTTPFVACIIPNHE 664
Qy 642 KMDPQFQAVVNLQRYSGMLFVTRIRKAGVAVRPPDFYKRYKVLNRLALPE---DV 698
Db 685 KRSGKIDAPLVLEQRLRCNGVLEBGRICRQGFPMRIVQFEPFRORYEILAN-AIPKGFMDG 743
Qy 699 RGCSTSLQLYDASNSEMOLGKTVPLRESLBOKEBREBEESHAMVIRAVLCPILAR 758
Db 744 KQACILMTKLEBDPMLYRIGOSKIFRTGVLAHLEBERDLKTTDVIMAFQACRGYLAR 803
Qy 759 KOYRK---VLVCVVIQKNYRAFLRRFLHKKAAIVFQKQRGQIARRVYRQL--LAE 813
Db 804 KAPAKQOQLTAKKVIQRCNCAVYL-----KLRNMOMWMLPFKVPVPLQ 846
Qy 814 KRQOEKKQOEKKEKKEKREBEREREREREAELRAQO--EETTRQOQLEALQSKQKA 871
Db 847 VTRQEBEMQAKEDLOKTKERQOKAENELKELOKISQLEEFKQLLOEQLOA-----ET 900
Qy 872 ELTRELEKOK-----ENKOVEEIL-----RLEKEIEDLOMKQOELSLTEASLOKLOER 921
Db 901 ELTAEBEBKVRILAACKQLEELTLEHMEALTEEBEDGQOLOAKRK-----KMAQOK 952
Qy 922 RDOELRLLEBACRAAQEFLSINPDEI--DECVNIERSLSGSEFSSLAESACEKRP 979
Db 953 LDLEBQOELEBAAR-----QXLOLEKVTAEKIKYLE-----984
Qy 980 NMFPSQPYREBEVDEGEADDDAFKOSPNSPSEHSHDQRTSGRTSDSDSEEDPYANDTV 1039
Db 985 -----DEILVMDQNNKLS--KERKLEBRISDL-TTNLAEBEERKAKNLTJK 1027
Qy 1040 VPTSPSADSTVLLAPSVQDSGLHNSSGESTYCPMPONAGDLPSPGDYVDYDDYEDGA 1099
Db 1028 LKRNK-----HEMNISELEVRLLKE-----EKSFQELE---1054
Qy 1100 ITSGSVTFNSYSGQSPDYRCVGTYNSSGAYRFSSSEGAQSSPFDESEEDPDSRF-DTD 1158
Db 1055 -----KLKRLKESDADDPHQIADLOAQIAELK 1082
Qy 1159 DELSTYRDSYSCVTLPEYHSLYMGGLMNSWKRRVCVLDKETFIMPSKQALQKQMW 1218
Db 1083 MQLAKKEBELQALAR-----LDDE-----IAQKNNALKK--1 1113
Qy 1219 HKKGGSSSTLSRNWKKRMFVLROSKLMPENDSEELKGTAVRTAKELIDNTTENG 1278
Db 1114 RELBGHISDLOEDLSEBAARKAERKOKRDLGELELAKTELE---DLDSTATQO-- 1166
Qy 1279 DIIMADRTFLIAESPEDASQMFVLSQVASTDOEIOEMHDEQANPONAVGLTDVGLID 1338
Db 1167 -----ELRAKREGEVTLKKAALDEBFRSHQAQVQEMRQNAQO-----1203
Qy 1339 SVGASDSPDRPNSFVITIANRVLAHCADPPEENHMTITLQNSKQDTRYVGEQFIYRGWL 1398
Db 1204 --AVELELTQLEQFKRAKAN--LDKRNKQTLKEXENADLA-----GELRYVGLQ-----A 1246
Qy 1399 HKEVKNSPKMSLKLKKRMFVLTHNSLDYVKSEKXNALKGLTVLNSLGSVYRPPDEKIRK 1458
Db 1247 KQVEHKKKKLEAQOVB-----LOSCKS--DGERARA 1276
Qy 1459 ETGYMNVTVYGRKHCYRLYTKLNEA-----TRMSVYQNVVTDTKAP 1500
Db 1277 EL---NDKYHKLQNEVESVTGMLNEABEKAIKLAKOVASISLOLQDTQLELDEFTQOKAN 1333
Qy 1501 IDTPTOQLIDIKENCCLNSDVVEQIYKRNPILRVTHPLHPLPLPYGDINLNLKDXG 1560
Db 1334 VSTFKRQLEBE--RNSLODQDLDEMEAKGNLEBH-----ISTNLQLOSDSK- 1377
Qy 1561 YTTLOD-----EAI-----KIFNSLOQLESMSDPIPIIGILQTOGHDLRPLRDELXY 1606
Db 1378 -KTLQDPASTVALEBGGKRFQKEIENLTQOEKAAAYDKLE--KTYNRLOQELDLIV 1433

```

```

Qy 1607 COLIKQTNKVPHPGSGVNLVSWOILTCLSTPELPSRGILKYLKFHLKRIREQPQTEMEK 1666
Db 1434 VLDNQROQLVSNLEKKQKRFD-QLLA-----BEKNISS--KYADERDRABAAREKKT 1483
Qy 1667 YALPTYESLK-----KTKREFPVSRDEIEALYIR-----1696
Db 1484 KALSLARALEALKEAKEBELERTNKMLKAMEDLVSSKQDVGVHLEKSKRALETOMBE 1543
Qy 1697 -----QEMSTVYCHGGGCKTITNSHTTAGEVVEKLIRGLAMEDSRN-----MFLAF 1744
Db 1544 MTKOLELEDELOATEDAKLRLVNMQALKGO-----FERDQARBQEBEKROLOROLH 1599
Qy 1745 EYNGHVDAKIESRTVADVLAKFEKLAATSEVGDLPWKFPYFLYCFLTDVNPVKOSVEPA 1804
Db 1600 EYTELEDERKQRALAAAKKLE-----GDL-----KO-----1628
Qy 1805 PMFEOAHEAVITHGHHPAPENIOVLAALRYLOQDYLTHAAIPLPEVYYSIQR-----1858
Db 1629 --LELOADSAIKGR---EBAIKQRLQAOQKDPQRELEBDARASRDEIFATAKENEKKA 1682
Qy 1859 --LKARISQSTKTPPCR-----LEKRR-----TSFLEGLTRRSFRFGSVYR 1899
Db 1683 KSLBADLMOLOEDDLAABERAKQADLEKEBELAEBLASSLSGRNALQDEKRLAARIAQLE 1742
Qy 1900 QKVEBEO-----MLDMWK-----EVSASARASIIDKRRKFGQMNQ--EQAMAKYMALIK 1947
Db 1743 BELEEBQGMEMASRVKATQQAQOQLNELATERSTQKNKSANQOJEBROKELRSKLH 1802
Qy 1948 EWPQ-----YGSTLDVECK 1962
Db 1803 EMEGAVKSKFKSTIAALEAK 1822

```

## RESULT 13

```

US-10-171-311-162
; Sequence 162, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 1972
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-171-311-162

```

## Query Match

```

12.4%; Score 1333; DB 9; Length 1972;
Best Local Similarity 24.2%; Pred. No. 1.4e-68;
Matches 514; Conservative 362; Mismatches 746; Indels 498; Gaps 72;

```

```

Qy 8 VWLRNNGHF--PSTVNSCAGIYVFRTDYGVQVFTYKQSTTTHQCVTMAHPNNEGCVDM 65
Db 36 VVVPSEKQGFPAASIKERKGEVVELVENG-----KQVYQKODIQKMNPKFSPKVEDM 90

```



QY 66 ASLTJLHGGSIWYNLFQRYKRNQIWTYIGSILASVNPYQPIAGLYEPATMEQYSRRHLGE 125  
 Db 91 AELTJCNLEASVLAHNRERYPSGLITYSGLFCVNVNPGHLP-IYSEKIVDMYKKGKRE 149  
 QY 126 LRPPIPALANEYCRLMKRHDNOCILIKESGAGTSESTKLILKPLSVISQOSLELSKE 185  
 Db 150 MPPHYAALADTYRMLDOREDOSILCTGESGAGTKENTKVIQYLAVALVASHKKGKOTS 209  
 QY 186 KTSCEBRIALBESSPIMEAFGNATVYNNNSRFGKVOLNICOKNIOGRIIVDCILS6Q 245  
 Db 210 ITGELKQOLQANPILIEAFGNATVYNNNSRFGKIRINFDVTYIVGANI-ETYLEK 268  
 QY 246 NRVNQNPEERNYHIFYALLAGLEHEBEREFLSTPENYLYNOSGVEDKTIISQESPR 305  
 Db 269 SRAIQARDERTFTHFYWMIAKAKEMRSDLLBEGNNYTPLS-NGFVPIPAQODDMQ 327  
 QY 306 EYITMDVWQPSKEKEVREVSLLAGILHGNIEF--ITAGAQVSFKTALGRSABLGL 362  
 Db 338 EYVEAMAIMGFSEEBOLSLIKYVSSVLOGNIVFKKERNTDQASMPDNTAAQKCHLMGI 387  
 QY 363 DPTOLT-DALTOBSMFLRGEILITPLNVOQAVD-SRDSLAMALYACCEPWVTKIINSRI- 419  
 Db 368 NITDTRSLITPRIXV--GRDVQKQKQTEQADPAVEALAKATYERLFRWILTRVAKAD 445  
 QY 420 -KGNEDFKSIGILDIGFENFEVNHFEQFNINYANEKLOHYENKHI FSLLEQLEYSREGIV 478  
 Db 446 KTHRGASFLIGLIDLAGFEIFEVNSPEQICINYTEKQQLFNHMFLEQEBYQREGIE 505  
 QY 479 WEDIMW-IDNGECDLIEK--KGLIALINEBHPQATDSTLBLEKHSQHANNHFFYK 534  
 Db 506 WAFIFGDLQPCIEILIERPNPFGVALLDEECWFPKATDKSPFEKCTEGESHPKFOK 565  
 QY 535 PRVAVN--NFGVYHAGEVOYDVRGILEKNRDPFDDILNLRBSPFDIYOLFPHVS-- 590  
 Db 566 PQOLDKTEFESIITHAGKVDYNASAWLTQMDPLDNVTSILNASSDKFVADLMKOVDR 625  
 QY 591 -----SRNODTLKCGSKHRR--PTVSSQPKVDSLHSIMATLSSSNFPFVACIKPMQ 641  
 Db 626 VGLDMQAKWTESSLPSASKTGKGMPTVQOLYK-BQLGKMTTINTNTNPFRCIIPNE 684  
 QY 642 KMPDPDQAVNLNOLRYSGMLETVRIRKAGIVRRPQD FYRYKYKLMNLALPE--DV 698  
 Db 665 KRSGLDLAFVLEQRCNGVLEGRICRQGFNRIYFOEFRQRYETILAN-AIPKGFMDG 743  
 QY 699 RGKTSILQVYDASNSEWQGLTKVFLRESLEQKLEKRESEVSHAAMVIRAHVGFAR 758  
 Db 744 KOACILMTALBELDPLVRYIGOSKIFERTGVLAHLEERDILKITDVIMAFQAMCRGYLAR 803  
 QY 759 KOYRK--VLYCVIIOKNYRAFLRLRRFLHLKKAIVQKQLRQIARVROL-LAE 813  
 Db 804 KAFARQOQLTMKVIQRYCAVY-----KLRMMQWRRLTTYKXPLIQ 846  
 QY 814 KREQEKKKQEBEKKGRBEERERERREARALPAQO--EETTRKOELALQSKQEA 871  
 Db 847 VTRQEEHQAKEDLOKTEROQKAKENELKELEQKHSQJTEKKNLLQEOQLA-----ET 900  
 QY 872 ELTRELEKQO-----ENKQVEEIL-----RLKEIJEIORMKEQOELSTIASLOKQER 921  
 Db 901 ELYABAEERKRYLAANKQOELIEILHEMEARLEEBEDRGQLOAERK-----KQAQOM 952  
 QY 922 RDOELRLLEEBACRAQOELIESINPEI--DECVENIERLSGSGSEFSSELAACEKRP 979  
 Db 953 LDLEQGLEEBEAR-----QKQLEKVTAAKIKKLE----- 984  
 QY 960 NFNFSQPIPEEVDGCFEADDAFKDQSPNSHSHSQDQTSIGIRTSDDSEEDPYMNDTV 1039  
 Db 965 -----DELVMDQNNKLS--KERKLEERISDL-TTNLALEEBEKAKANLTK 1027  
 QY 1040 VPTSPASDSTVLAPSVQDSGLHNSSGSESTYCMQONAGDLPSPDGDVDYQDDYEDGA 1099  
 Db 1028 LKNK-----HSMISELEVRLLKG-----EKSRQELE--- 1054

QY 1100 ITSGSVTFNSYGSQWSPDYRCVGTYNSSGAYRPSSEGAQSSFEDEEDPDSRF-DTD 1158  
 Db 1055 -----KLRKLEGGASDPHEQIADLOQIARILK 1082  
 QY 1159 DELSVRDSVYSCVTLPIFHSFLYKGLMSWKRRKCVLQDETMLFRSRQOALQOML 1218  
 Db 1083 MQLAKKEBELQALLAR-----LDDE--IAQRNNAKK--I 1113  
 QY 1219 HKGGGSSTLRRBMKKRWFLRQSKLWYFENDESEKLGTVETAKETIINTTKENGI 1278  
 Db 1114 RELEGHITDQEDDSEBAANAKKQKRDGEBELAKTLE-----DTLDSITQO-- 1166  
 QY 1279 DIIMADRTFHLIABSPDASQWFSVLSQVHASTDQEIQEMHDEQANFQNAVGLDVLID 1338  
 Db 1167 -----ELRAREQGVTVLKKALDEBTRSHBAQVEMQKQHA----- 1203  
 QY 1339 SVCASDSDPRNSPITIANRVLHONADTPREMHMITLORSKQDTRVBOGFIVRGML 1398  
 Db 1204 --AVELETOEOEFRAKAN--LDKNKQTLKEKNADLA-----GELRVLQO--A 1246  
 QY 1399 HKEVNSPMSLSKLKKRWFLTHNSLDYKSSSEKNALKGLTVLNSICSVVPPDEKIFK 1458  
 Db 1247 KOEVEHKKKLEAOVQO-----LOSKCS--DGERARA 1276  
 QY 1459 ETGVNNTVYGRKHCYRLYTKLNEA-----TRMSSVIONVTDTRAP 1500  
 Db 1277 EL--NDKVHLQNEVESVLTGMLNEABGKAIKLAQVASSSQLODQBELQOEBTROKLN 1333  
 QY 1501 IDTPQOQIIDIKENCLNSDVVEQIYKRNPLRTYTHNLHSLPLRPIYGDINMLKXG 1560  
 Db 1334 VSTKRLQLEER--RNSLDQDDEMEAKQNERH-----ISTNIOUSDSEK- 1377  
 QY 1561 YTTLOD-----EAI-----KIFNSLOLESMSDPIPIIOGILQTHDRLPRDEY 1606  
 Db 1378 -KLQDPASTYBALLEGGKRPQKETEMLTQOYEEKAAVYDLB--KTRNLOQSLDVL 1433  
 QY 1607 QOLIKQTRKVPHPGSVGNLYSMOILYTCLSCFPLSRGILKYLKRIARQOPGTEMEX 1666  
 Db 1434 VDLNQOQLVSNLEKKQKFD-QLLA-----BEKNIS--KVADBRDRAABAREKET 1483  
 QY 1667 YALFTYELK-----KTKCEPFPSPREIETALIR- 1696  
 Db 1484 KALSLARLLEBALBEKELBERTNKALKAMEDDLVSXKDVKNVHELEKSKRALETOMBE 1543  
 QY 1697 -----QEMTSVYHCGGSSCKITINSHTTAGEVEKILRGAMEDSN-----MPALF 1744  
 Db 1544 MKTQLEBELQATEDAKTLEVNMOALQO--FERDLOARDEJNEBKROLQORQLH 1599  
 QY 1745 EYNGHVDKALIESRTVAVDVLAKFEKLAATSEVGDLPWKFFYKLYCFLDTDVVPKDSVEFA 1804  
 Db 1600 EYETLEDERKQRLAALAAKKKLE-----GDL-----XD----- 1628  
 QY 1805 FMFQOAEAVYHGHHPABEENLOVLAALRLOYLOGDYTLHAALPBLEVYSLOK----- 1858  
 Db 1629 --LELOADSAIKGR--EBAIKQRLKLOAQOMKDFORELEBARARSRDIPATAXENEXKA 1682  
 QY 1859 --LKARISQSTKFTPCER-----LEKRR-----TSFLEGLTRSRFRGSVVR 1899  
 Db 1693 KSLBOLDQLOEDLAAERARKQADLEKEBLAELBASSGRNALQDEKRRLEARIQOE 1742  
 QY 1900 QKVEEQ-----MLDMWIK-----EYVSSARASIIDKRRKFOGMNQ--EQAMAKYALIK 1947  
 Db 1743 EELBEEQOMEMASDRYKATQOQJLSNELATERSSTQKXBSARQJLERQNKELRSKLN 1802  
 QY 1948 EMPG-----YGSTLPDVECK 1962  
 Db 1803 EMEGAVKSKFKSTIAALEAK 1822

RESULT 14  
 US-10-044-303-1  
 ; Sequence 1, Application US/10044303  
 ; Patent No. US20020137161A1

/ GENERAL INFORMATION:  
/ APPLICANT: Max-Planck-Gesellschaft e.V.  
/ TITLE OF INVENTION: Protein expression and structure solution using  
/ FILE OF INVENTION: Specific fusion vectors  
/ FILE REFERENCE: ST010209-EPA  
/ CURRENT APPLICATION NUMBER: US/10/044,303  
/ NUMBER OF SEQ ID NOS: 3  
/ SOFTWARE: Patent In Ver. 2.1  
/ SEQ ID NO 1  
/ LENGTH: 765  
/ TYPE: PRT  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: Partial myosin  
/ OTHER INFORMATION: sequence of Dictyostelium; Component (1) of the  
/ OTHER INFORMATION: recombinant protein M761-2R R238E  
US-10-044-303-1

Query Match 12.4%; Score 1329; DB 12; Length 765;  
Best Local Similarity 40.1%; Pred. No. 6; 8e-69;

Matches 299; Conservative 160; Mismatches 253; Indels 34; Gaps 16;

QY 8 WMLRENGQHFPPSTVNSCAGIIVFTDYQGVFTYK-----OSTTHQKVTAMHPTNEEGV 62

DB 39 IWYNPD-----PKERDYECEGEIVSET--SDSFFFTKVDQDQKDDANORNPFIKFDGV 92

QY 63 DDMASLTSLHGGSIWNLFORYKRNQIMTYIGSLASVNPQPIAGIYEPATMEQYSRRH 122

DB 93 EDMSLSYLNBPAPVFNLRVYINQDILYISGLPLVAVNPFKIP-ITYQEMNDYIFKGR 151

QY 123 LGLPRLPFAIANECYRCIMKRDNOCLIKGSGAGKTESTLILKFLSVISQOGLSLS 182

DB 152 RHEVAHFALISVAVARSMLDQRQNSLLITGSGAGKTEMTKVIVYLASVAGR-----N 207

QY 183 LKEKTSICVERALIESPPIEAFGNATVYNNSSRFGKVQVQNTICQKNIQGGRIYDCL 242

DB 208 QANGSVLLEQOILQANPILFAFGNATVYNNSSRFGKFIETIQNSAGFISGASIOSYLL 267

QY 243 SSQNRVVRNPGERNHYIFVYLLAGLHEHREBEFYLTSPBNHYLNQSGVEDKTI SDPE 302

DB 268 -EKSRRVPSSETERNHYIFVYLLAGLHEHREBEFYLTSPBNHYLNQSGVEDKTI SDPE 326

QY 303 SFREVITAMDVMQFSKEEVRVSRLLAGILHLGNIEFIT-AG-GAQSFKTALGRSAELL 360

DB 327 EFKITRQAMDIVGFSQEGQMSIFKTIAGILHLGNIEFKGAGGAVLXKXTALNASTVF 386

QY 361 GLDPTOLTALTORSMFLAGEELITPLNQAVDSQSLAMALYACCFEWTIKINSRLK 420

DB 387 GNPSTLEKALMPRIAGRDVAQHLNVEKSSSSRDALVKALYGRFLFMLVYKIKINVL 446

QY 421 GNEDFSIGILDI FGFEFNVNHFQFNINYNANEKIOEYFNKIFISLEDEYREGVLWE 480

DB 447 QERKAVFTGLVDSGEIFRQNSFEQCLNYTNEKIQOFPNHHMFLDEBEYLUKEKINT 506

QY 481 DIDW-IDNEGCDLIEKCL--GLALINEESHFPQATDSTLLEKLSQHA--NNHFPVKR 536

DB 507 FIDFGDSQATDILIDGRQPGILALDEQSVPRNATDNTLITKLHSHSKNAKKEER 566

QY 537 VAVNNGVHYAGEVOYDVAGILEKRDTPFRDDLNLNBSRFDIYDLF--HSSRN 594

DB 567 FSKTEGVTYHAGQVWEYIOMLEKNKDPLOQDLIECFQSDSDNVVTKLFPNPIASR-- 624

QY 595 QPTLKSGHRRFTVSOPFVDSLSHLMATLSSNFFVRCIKPNQKMPQDQAVLN 654

DB 625 ---AKGANP--ITVAQYK-EQLASLMKLTSTNHPVRCIIPNNKQUPAKLEDEVRVD 678

QY 655 QURSGMLTETVIRKAGVAVRFPQFYRYKYVLMENLAL-PEDVAGKTSLLQTDASN 713

DB 679 QURNGVLEGRITRKGFPRIIYADFVIRYLLANVPDADSDQKATDAVLKHLNIDP 738

QY 714 SEMQAGKTVFPLRESLEQLEKREER 739

DB 739 EQYREGITKIFFRAGQILARIPEARQ 764

RESULT 15

US-09-927-597-2

/ Sequence 2, Application US/09927597

/ Publication No. US20030032018A1

/ GENERAL INFORMATION:

/ APPLICANT: Malik, Fady

/ APPLICANT: Berard, Christophe

/ APPLICANT: Freedman, Richard

/ APPLICANT: Craven, Andrew

/ APPLICANT: Sakowicz, Roman

/ APPLICANT: Hartman, James

/ TITLE OF INVENTION: Human smooth muscle myosin heavy chain

/ FILE REFERENCE: CYTOPO18

/ CURRENT APPLICATION NUMBER: US/09/927,597

/ NUMBER OF SEQ ID NOS: 17

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 2

/ LENGTH: 1945

/ TYPE: PRT

/ ORGANISM: Human

US-09-927-597-2

Query Match 12.4%; Score 1328.5; DB 9; Length 1945;  
Best Local Similarity 24.2%; Pred. No. 2; 4e-68;

Matches 515; Conservative 365; Mismatches 742; Indels 505; Gaps 74;

QY 8 WMLRENGQHF--PSTVNSCAGIIVFTDYQGVFTYKOSTTHQKVTAMHPTNEEGVDM 65

DB 36 VWVPSKQGFEEAASIKKEKGEVVELVENG-----KVTYQKODIQKNNPFPKSKVDM 90

QY 66 ASLTSLHGGSIMYNLFORYKRNQIMTYIGSLASVNPQPIAGIYEPATMEQYSRRHIG 125

DB 91 AELTLNBSAVLNHREYFSGILTYISGLFCVNVNPKHLR-ITYSEKIVDYKGRKHE 149

QY 126 LPPHIFAIANECYRCIMKRDNOCLIKGSGAGKTESTLILKFLSVISQOGLSLS 182

DB 150 MPPHIALADRYARRMLDQREQSLCTGSGAGKTEMTKVIVYLASVASHKQKOTS 209

QY 183 LKEKTSIC---VERALIESPPIEAFGNATVYNNSSRFGKFIETIQNSAGFISGASIOSYLL 238

DB 210 ITQGSFAYGELKQLQANPILFAFGNATVYNNSSRFGKFIETIQNSAGFISGASIOSYLL 268

QY 229 DCILSSQNRVVRNPGERNHYIFVYLLAGLHEHREBEFYLTSPBNHYLNQSGVEDKTI 298

DB 269 EYLLKSRALRQARDERTFHI FYMIGAKRKMSDLLLGFFNNYFLS-NGFVPPIPA 327

QY 299 SPOESFREVTAMDVMQFSKEEVRVSRLLAGILHLGNIEFIT--ITAGGAVSFYALGR 355

DB 328 QDDMFQETVERMAIMGSEBEQSLIKVSVYDQGNIVFKKENTQOASRPDNTAK 387

QY 356 SAEILGLDPTOLT-DALTORSNFTLAGEELITPLNQAVD--SRDSLAMALYACCFEWIK 413

DB 388 VCHLGINVDTPTSRILTPRIKIV--GRDVQQAQFQEDADFAVEALATATERLERLWILT 445

QY 414 KINSRI--KGNEDPFSIGILDI FGFEFNVNHFQFNINYNANEKIOEYFNKIFISLEDEY 471

DB 446 RVNKLADKTHRGASFLDILDIAGEIFEVNSFELCLNYTNEKIQOFPNHHMFLDEBE 505

QY 472 YSRGLVMEIDW-IDNEGCDLIEKCL--KGLALINEESHFPQATDSTLLEKLSQHA 527

DB 506 YOREGIEWNFIDFGLDQPCIEILFRPNPPGVALLDBECMFPAATDSFVEKCTDQ 565

QY 528 NNHFPVKRPAVN--NFGVKAHYAGEVOYDVAGILEKRDTPFRDDLNLNBSRFDIYDL 585

DB 566 SHPKQFKQKQDKTEFSIHYAGVDYNASAMLTQNDPLINDVNTSLNASSDQKFAVDL 625

QY 566 FEHVS-----SKNNDPTLKSGSKRR--PTYSQPKVDSLSHLMATLSSNFPFR 634

DB 626 WQDVRIVLDQMAKMTSSLPASAKTKKGMFTVGOLYK-EQLKMLTTLNTPNFVR 684

Oy		635	C I K P M Q M G P O F D A V U L N O R S Y S G M E T Y R I R A G A V A N R P P D F X K A K Y L M B N I A L	694
Db		685	C I I P H E R S G S K L D F V L E O J R C G V L E G I R C I O G P R N N I V F E E F O R Y E I L A A N - A I	743
Oy		695	P E --- D V A R G K T S L O L Y D A S N S E W O L A K T I V F L E S E I O K L E K R E E V E H A A M V T R A H	751
Db		744	P K G F M D G Q A C I L M I K A L E D R P N V R I G O S K I F P T G V L A N H E E R D L K I D V I M A P Q M	803
Oy		752	V L G F L A R Q O Y K -- V L Y C V I I O Q N Y A F L I R R A F L H L K K A I V F Q O L G O L A R P Y R	808
Db		804	C R G Y L A R A F A F R Q O O L T A M K V I O R N C A Y L ----- K L N W O M W R L F T	846
Oy		809	Q L -- L A E K R D E E K K O E E E K K K E E E R E R E R E R E R A E L R A O Q -- E E T R K O O L E A L	864
Db		847	K V P L O V T R D E E E M Q A K E D E L O K T K E Q O A R E N L K L E O G N S L T E K Y L B O Q L O A -	905
Oy		865	O K S O K E A L T E L E K O K ---- E N K O V E I L ---- F L E K E I D I O R M K E O Q E L S L T E A S	914
Db		906	----- E T E L V A E A E M R V L A K O K E L E I I I H E M A R E L E E E D R Q O L A R K -----	953
Oy		915	L O K I O E R D O E L R L E E E A C R A A O F L E S I N F D E I - D E C V R N I E R S L G S E S P S E L A E	972
Db		954	- K M A O M I D L B O L E E E E A R ---- O K L O E K T A A K I K G E -----	991
Oy		973	S A C E K P N F N S O P A P E E E V D G E F A D D A F O S P N E P E H G S D O R T S I G I R T S D S S E E D	1032
Db		992	----- D E I V M D Q O N K L S -- K E R K L B E R I D L - T T N I A E E E	1072
Oy		1033	P Y M D T V P T S B S A D T V L L A P S V O D S G L S N S S G S T Y C M P O N A G D L P S D G D Y D Q	1092
Db		1028	K A K U L T Y K N K ---- H E S M I S E L V L R K E ---- E K S R	1057
Oy		1093	D O Y E D G A I T S G S Y T F S N S Y G O M S P D Y C S V G T Y N S G A V R F S E G A Q S P E D E E P F D	1152
Db		1058	Q E L E ---- K L K R L E G A S D P H E I A D L Q	1082
Oy		1153	S R F - D T D E L S Y R D S V Y S C T L P Y F H S F L Y M K G L M S W K R M C V L D E T F L M R S K O E	1211
Db		1083	A O I E L K Q L A K K E E L O A L A L R ---- L D D E ---- I A Q N N	1115
Oy		1212	A L K G O M L K K G G S S T L S R M M K K M F W L B O S K L M F E N D S E K L K G V E R T A K E I I D N	1271
Db		1116	A L K R - I E L S G H I S D L O E D L D S E A R N K K E K O R D I G E L E A K T E L E ---- D T L D S	1168
Oy		1272	T T K E N G I D I M A D R T F H L I A S P E D A S O F S V L S C V N A S T O E L O E M H D E O A N P O N A V G T	1331
Db		1169	T A T Q O ---- E L R A K R E G E V T V L K A L D E T S H A Q V E M O Q K N A Q	1210
Oy		1332	L D V G L I D S V C A S D E P R N S F U T I T A N R V L H C M A D T P R E M I N H I T L L O R S K G D T R V E G O E	1391
Db		1211	---- A V E E L T B O L E O F R A K A N - L D K N O K T L E K E N A D L A ---- G E L R V I G O -	1252
Oy		1392	F I V R G M L K E V A N S P R M S L K K K M F V L T H N S L D Y K S S E K M A L K G T V L N I N S C V V P	1451
Db		1253	---- A O E V E H K K K L E A Q V O E ---- L O S C S --	1276
Oy		1452	P D E K I F K E T G Y W N V T V Y R K H C Y R L Y L K L N E A ---- T R M S S V I O N	1493
Db		1277	D G E R A R A E L -- N D K V H K L O N E V E S V T G M L N E B A K I K A L D V A S L S Q L O D T O E L L O E	1333
Oy		1494	V T D T A R I D T P T O O L I O I K E N C S N D V E O L Y K N P L R T I T H N L H S P L R P R G D I N L	1553
Db		1334	E T R O L A N S T K R O L E E E - R N S L O D O Q D E E A E A Q N L E R H ---- I S T N I	1378
Oy		1554	N L K D K G Y T T I O D ---- E A I ---- K I F N S I O O L E S G S D P I P I O G I L O T G H D L R	1599
Db		1379	Q L S D E K - K K L O D P A S T V A L E B E G K R F O K E I E N L T O Q Y E K A A V D L E -- K T X N R L Q	1433
Oy		1600	P L R D E L Y O Q L I Q T N K V P H P S G V N L Y S M O I I T C S T F L P S R G I L K T L K N L K R I R E O F	1655
Db		1434	Q E L D D L V D L D N O R O L V N L E K O R K F - Q L A ---- B E K I N S -- K A D E D R A E A	1483

```

0Y 1660 PGTEMEKXALPTYSLSK-----KTKCRFPVSPSDEI:EA:LI:HR----- 1696
Db 1484 EAREKETKALSLAALEALEAKELELERTNKMLKAEMEDLVSS:KDDVGKXVHELEKSKRA 15433
0Y 1697 -----QEMTSTVYCHGGGSCKITINSHTTAGVEVEKLI:RGLAMEDSRN----- 17399
Db 1544 LETQMEEMKQTLEBELDELQATDEDAKURLFLENNQALGQ-----PERDLOARDEQNEBEKRR 159399
0Y 1740 --MALPEYNGHVDKAIESTRTVADVLAKPEEKLAAITSEVGDDLPMKFFYKLYCFLDTDNVP 17977
Db 1600 QLOQOLHEYELELDEDERKORALAAAKKLE-----GDL----- 16333
0Y 1798 KDSVEPFAFMPEQAEVAVIHHGHPAREENLQALALRYLOVQGDVTLHAATPRLEEVYSLQ 1857
Db 1634 KD-----LELOQDSAIKGR---EBAIKQLRKLQLOMKDFORELEDAARSDEIFATA 1682
0Y 1858 R-----LKARISOSTKTTFTPCR-----LEKRR-----TSFLTGLTRRSF 18932
Db 1683 KENEKKAKSLEADLMQOEDPLAAERARKQADLEKELELABELASLSGRNALQOBKRRLB 17422
0Y 1893 RTGSVVRQKYEEO-----MIDMKIK-----EEVSSAPASTIDKRRKFOQGNQ--EQMA 1940
Db 1743 ARIQOLEBELLEEOGNMEAMSDRKRKATQOAEQLSNELATERSTQOKRNESARQOLERONK 18020
0Y 1941 KYMALIKEMPS-----YGSTLFDVECK 1962
Db 1803 ELRKSLHEMGAVAKSKFRKSTJALAEAK 1829

```

Search completed: July 14, 2003, 18:27:25  
Job time : 62.2242 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 6.99422 Seconds  
(without alignments)  
2620.804 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3252  
Sequence: 1 MSTDQEROINTEYAVSLLEQ.....STGTEEPEDLGEWALPPV 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6CTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfill1.pep:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516.5	15.7	589	2	US-08-317-305-4 Sequence 4, Appl
2	516.5	15.7	589	3	US-08-862-508-4 Sequence 4, Appl
3	516.5	15.7	589	2	PCT-US95-12508-4 Sequence 4, Appl
4	511.5	15.5	589	3	US-08-317-305-2 Sequence 2, Appl
5	511.5	15.5	589	3	US-08-862-508-2 Sequence 2, Appl
6	511.5	15.5	589	5	PCT-US95-12508-2 Sequence 2, Appl
7	422.5	12.8	516	4	US-09-154-750A-86 Sequence 86, Appl
8	357.5	10.9	642	4	US-09-337-307A-3 Sequence 3, Appl
9	357.5	10.9	642	4	US-09-337-307A-3 Sequence 4, Appl
10	329	10.0	553	4	US-09-068-655-6 Sequence 6, Appl
11	309.5	9.4	530	1	US-08-307-499-29 Sequence 29, Appl
12	309.5	9.4	530	4	US-09-299-268-29 Sequence 29, Appl
13	234.5	7.1	500	1	US-08-307-499-20 Sequence 20, Appl
14	234.5	7.1	500	4	US-09-299-268-20 Sequence 20, Appl
15	195.5	5.9	478	4	US-09-137-223A-2 Sequence 2, Appl
16	192.5	5.8	109	4	US-08-553-541B-7 Sequence 7, Appl
17	192.5	5.8	109	4	US-09-268-202-7 Sequence 7, Appl
18	167.5	5.1	803	4	US-09-063-035-2 Sequence 2, Appl
19	135.5	4.1	706	1	US-08-074-967-2 Sequence 2, Appl
20	135.5	4.1	706	2	US-08-553-541B-2 Sequence 2, Appl
21	135.5	4.1	706	4	US-09-468-202-2 Sequence 2, Appl
22	135.5	4.1	706	5	PCT-US94-06669-2 Sequence 2, Appl
23	125.5	3.8	104	1	US-08-340-203A-6 Sequence 6, Appl
24	125.5	3.8	104	2	US-08-452-427-6 Sequence 6, Appl
25	125.5	3.8	104	3	US-09-085-407-6 Sequence 6, Appl
26	125	3.8	106	1	US-08-340-203A-10 Sequence 10, Appl
27	125	3.8	106	2	US-08-553-541B-5 Sequence 5, Appl

28	125	3.8	106	2	US-08-452-427-10 Sequence 10, Appl
29	125	3.8	106	3	US-09-085-407-10 Sequence 10, Appl
30	125	3.8	106	4	US-09-268-202-5 Sequence 9, Appl
31	125	3.8	110	2	US-08-553-541B-9 Sequence 9, Appl
32	125	3.8	110	4	US-09-268-202-9 Sequence 3, Appl
33	124.5	3.8	569	4	US-09-137-223A-3 Sequence 6, Appl
34	123.5	3.8	108	2	US-08-553-541B-6 Sequence 6, Appl
35	123.5	3.8	108	4	US-09-268-202-6 Sequence 6, Appl
36	121.5	3.7	549	4	US-09-245-041-9 Sequence 2, Appl
37	121.5	3.7	1260	4	US-09-245-041-2 Sequence 7, Appl
38	121	3.7	104	1	US-08-340-203A-7 Sequence 7, Appl
39	121	3.7	104	2	US-08-452-427-7 Sequence 7, Appl
40	121	3.7	104	3	US-09-085-407-7 Sequence 7, Appl
41	120.5	3.7	1198	4	US-09-245-041-131 Sequence 131, App
42	120.5	3.7	1198	4	US-09-794-236-3 Sequence 3, Appl
43	120.5	3.7	1429	4	US-09-245-041-130 Sequence 130, App
44	120.5	3.7	2787	4	US-09-245-041-15 Sequence 15, Appl
45	115	3.5	105	1	US-08-340-203A-8 Sequence 8, Appl

## ALIGNMENTS

RESULT 1  
US-08-317-305-4  
; Sequence 4, Application US/08317305  
; Patent No. 5863744  
; GENERAL INFORMATION:  
; APPLICANT: Avraham, Shalom  
; APPLICANT: Avraham, Hava  
; APPLICANT: Geopman, Jerome E.  
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
; TITLE OF INVENTION: SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
; STREET: Ten Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,305  
; FILING DATE: 03-OCT-1994  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Ph.D., Kathleen A.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: DH-001XX  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-2290  
; TELEFAX: (617) 451-0313  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-317-305-4  
  
Query Match 15.7%; Score 516.5; DB 2; Length 589;  
Best Local Similarity 25.5%; Pred. No. 1.3e-44;  
Matches 155; Conservative 94; Mismatches 237; Indels 121; Gaps 17;  
  
CY 11 TTYAIVLEQLKFLFEQQLFTDIVIVIGETBPCPKMTLACSSYFRPMFMSGLSESXKT 70  
DB 26 SSYADSVLTHNLRLROORLFTDVLHLHAGNRTFPCRAVLACSRYPFAMFSGGLSESODS 85

```

0Y 71 VHLELN-VDAATLQIITAYTGNLMMNSTYQLJETAECFLOVEVLORCREYLLKKN 129
Db 86 EVNFDNISHPEVLELIDYASISRVIIINEENNEESLEAGDMLEFODIRDACAFLEKKN 145
0Y 130 AENCYLLSPADLFSCEEELKOSAKRWVEHKFTAVYQDAPMOLSHLLIDILSSDLNVE 189
Db 146 PTNCGMLLSDAHQCTKYLKSWRCLSNFOIRKNEDPDLQPDMMVYVGLSSFELETE 205
0Y 190 KEETVREANMLLEINTBSRSQYLSVLSQIRIDALSEVTORAMPGLPNDKSVVYVGL 249
Db 206 DERLYVESAMNNSYDLKRYCYLPBELLQTVRLALPAIYLM-ENVAMBEL 255
0Y 250 YSMRPFKPRJRMGTKE--EMMIFLJESSENPECLSVSVCSYSPQ-----A 292
Db 256 ITKORK-----SREIYEAIRCLKTLONDGVTSILCARBKTHGHALFLIGGOTPMC 307
0Y 293 EKVY-----KLCSPPADL---HKVGTVVTDPNDIYIAGQVYLNKNTKTHSTKSLQTA 343
Db 308 DKLYLVDOKAKIIPKADIPSPKRESACALGCKVYITGG-----RSENGVSD----- 357
0Y 344 FRTVNCFYW-FDAQONTWFPKTPMLFVKRIPSLVCEGYIYALGDS-----VGGELN 395
Db 358 -----VWVYDTLHEBMSKAPMLVARFGHSGABELKGLCYVVGHTATGCLPASPSVS 410
0Y 366 RRTVERDYDEKCBMTNVSLPCAMQMSAAV-----VANDCIYMTNLNMYCY 442
Db 411 LKQVBOYDPTTKMTMVAPLRBGVSNAAYASAKLKLPAFGSTVSHD-----KLPRVOCY 465
0Y 443 FPRBSWVMARQTSRSPASAAAAGFKIFVYIGLHATNSGIRLPSGTVDGSSVTEIY 502
Db 466 DQCEMRKSPACPCPOKRYTAAAVLGNQIFIM-----GDLEFSACAYKE 511
0Y 503 DVNKNEMKMAANI PAKRYSDPCVRAYVINSLSLCEFNRETHLBRAXYTYQ-----YD 555
Db 512 NSETYQMTKVGVDYAKRMS---CHAVASGKLYV-----GGVFGIQRCKTLDYD 559
0Y 556 LEIDRMS 562
Db 560 PTLIDVNN 566

```

RESULT 2  
 US-08-862-508-4  
 Sequence 4, Application US/08862508  
 Patent No. 606451  
 GENERAL INFORMATION:  
 APPLICANT: Avraham, Shalom  
 APPLICANT: Avraham, Hava  
 APPLICANT: Groopman, Jerome E.  
 TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lalive & Cockfield, LLP  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/862,508  
 FILING DATE: 23-MAY-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/317,305  
 FILING DATE: 03-OCT-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Silveri, Jean M.  
 REGISTRATION NUMBER: 39,030  
 REFERENCE/DOCKET NUMBER: NER-259DV

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-6214
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-862-508-4

```

Query Match	15.7%	Score 516.5	DB 3	Length 589
Best Local Similarity	25.5%	Pred. No. 1.3e-44		
Matches 155; Conservative	94	Mismatches 237	Indels 121	Gaps 17

```

0Y 11 TEYAYSLLOQLKFEEOOLFETDVLIVYAGTEFPGCHKVLLTCSSEYFPAMMSGISESKOT 70
Db 26 SSIYDSVLTHLNLLOQRLFTDVLVLAHQNRFFPCHRAVLAACSYFFPAMSGSKESQDS 85
0Y 71 HYLHRL-VDAATLQIIITYAYATGNLAMDSTVEOLYEFACLOVEDVLOHCREYLKIKIN 129
Db 86 EVAFDPSIHPEVLELLDIAVYASSRVITINEBNAESILLEGDMLEFODIRDACAFLEKXNH 145
0Y 130 AENCVALLSFADLFSCSEELKOSAKRMVNHKRTAYYHODAFMQOLSHDILLIDISDNLANE 189
Db 146 PTNCLMELLSDAHOQTKLYELSMCMUSNQTQIRKNEDFLOQRPQDMMVOLLSSSELETE 205
0Y 190 KEETVEAAMLELVYNTESRSQYLSVLSQIRIDALSEVTQRAWFQGLPENNDSVVVQGL 249
Db 206 DEBLVYESAMMISYDLKKRYCYLPBELLQYRLALMLPAIYLM-----ENVAMBEL 255
0Y 250 YSMPEFFPKRLGMYKE-EMMIFLEASNEPCLYSVCYSPO-----A 292
Db 256 ITRKOR-----SKEIYEBAIRCKLILONDVWVSLCARPKTGHALFLLGQTFMC 307
0Y 293 EKYV-----KLCSPPADL---HKUGYVWPDPNDIYIAGGOVPLKONTKTHSKTSKLOTA 343
Db 308 DKLIYLDQAKKIIIPKADIPSPRFEFSCAIGCKVYITGG-----RSGENGVSKD----- 357
0Y 344 FRTVNCFYV-FDAQONTWPKTIPMLFVRIKPSLVCBGYIYAIJGDS-----VGGELN 395
Db 358 -----VWVYDTLHEBMSKAAPMLVAFHGHSABLKHCLYVVGHTPAATGCLPASPSVS 410
0Y 396 KRIVEYEDTEKKEBMTMVSPLPCAMOMSAV-----VHDCIYVMTLNLMTCY 442
Db 411 LKQVEYEDPTTKMKTMVAPLREGVSNAAVSAKULPAFGTGSVSH-----KLPRQOC 465
0Y 443 FPRSDWVEMAKRQTSRSPASAAAFGDKIFYIGGLHATNSGIRLPFGTVDGSSVTEIY 502
Db 466 DQENMNSPAPCCPQPMRYTAAVUNGINOIFMG-----GDTEFBSACSNYKF 511
0Y 503 DVNKMEKMAANIIPAKRYSDDPCVRAVVISNLSGVMEETHLNEBAKYVTVQ-----YD 555
Db 512 NSETYQMTKGVDTAKRMS---CHAVASGNKLYV-----GGFYGIQRCCKTLCYD 559
0Y 556 LELDRMS 562
Db 560 PTLDVWN 566

```

```

RESULT 3
PCT-US95-12508-4
; Sequence 4, Application PC/TUS9512508
; GENERAL INFORMATION:
; APPLICANT: Avraham, Shalom
; APPLICANT: Avraham, Hava
; APPLICANT: Groopman, Jerome E.
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
;

```

STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/12508  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-12508-4

Query Match 15.7%; Score 516.5; DB 5; Length 589;  
Best Local Similarity 25.5%; Pred. No. 1.3e-44;  
Matches 155; Conservative 94; Mismatches 237; Indels 121; Gaps 17;

11 TEVAVSLEQLKFLFEOQLFTDVLIVEGTEPFCNKVLTATSSYFRAMFSGLSSEKOT 70  
26 SSYADSVLTHLNLNQRLFTDVLHAGNRTPFCHRAVLAACSRFEAMFSGLSSESODS 85  
71 HVLHNL-VDAATLQIIITFAATGNLAMDSTVEQLYETACFLQVEDVLQRCREYLIIKIN 129  
86 EVNPNDSIHPEVLLELLDYAISRVINENNESLLEAGDMLEFQDIRACAEFLKINH 145  
130 AENCVRLLSPADLFSCCEELKOSAKRVEHKFTAAYVHQAFLMQLSHDLLIDILSSDNLYE 189  
146 PNCGLMLLSDAHQCTKLVELSWRMCLSNFQIRKNEDELQLPQDMVVOQLSSSELEFE 205  
190 KEETREAAAMLLENTSRSQYLSVLSQIRIDALSTYQRAWQGLPNDKSVVVOQL 249  
206 DERLYESAMNWSYDLKRRYCYLPQLQTVRLALLPAILYM-----ENVAMEBL 255  
250 YKSMKPFKPRLGMTKE--EMMIFIASSENPCSLYSVCYSPQ-----A 292  
256 ITRKOR-----SKEIYEAIRCKLKLQNDGVVTSICARPKRTGHALFLGGOTFNC 307  
293 EKVY-----KLCSPADL--HKVGTVPDNDIYIAGQVPLKNTKTNHSTKSLQTA 343  
308 DKLVLVDQAKKEIIPKADIPSPKEFSACALGCKVITG---RSENGVSKD----- 357  
344 FRTVNGFYV-PDAQQNTWFPKTPMLFVRIRKPSLVCESEYIYAGDS-----VGEELN 395  
358 -----VVVYDPLHEEWSKAAPMLVARFGHGSALHCLKYVGGHTAATGCLPASPVS 410  
396 RRTVERYTEKDEMTWVSPLPACAMQWSAAV-----VVHDCIYVMTLNLNMYCY 442  
411 LKQVEYDPTTKMTKTVAPLREGVSNAAVVSATLKLFAFGGTSVSHD-----KLPRVQCY 465  
443 PRPSDSWEAMAKQTSRSPASAAAFGDKIPIYIGGLHIATNSGIRLPSGTVDGSSVTEIY 502  
466 DDCENKMSVPATCPQPMRYTAAVLAQLNGQIFIMG-----GDTFSACGAYKF 511  
503 DVNKNQKMAANIPARKYSDPCYRAVVISNLCVPEKEETHLNBRAKYVVO-----YD 555  
512 NSETYQWTVYGVDTAKRMS--CHAVASGNKLYV-----GGYFGIQKCKTILDYD 559

QY 556 IELDRLWS 562  
DB 560 PTLDVWN 566

RESULT 4  
US-08-317-305-2  
Sequence 2, Application US/08317305  
Patent No. 5853744  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
TITLE OF INVENTION: SAME  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,305  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DR-001XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-305-2

Query Match 15.5%; Score 511.5; DB 2; Length 589;  
Best Local Similarity 25.4%; Pred. No. 4.3e-44;  
Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

11 TEVAVSLEQLKFLFEOQLFTDVLIVEGTEPFCNKVLTATSSYFRAMFSGLSSEKOT 70  
26 SSYADSVLTHLNLNQRLFTDVLHAGNRTPFCHRAVLAACSRFEAMFSGLSSESODS 85  
71 HVLHNL-VDAATLQIIITFAATGNLAMDSTVEQLYETACFLQVEDVLQRCREYLIIKIN 129  
86 EVNPNDSIHPEVLLELLDYAISRVINENNESLLEAGDMLEFQDIRACAEFLKINH 145  
130 AENCVRLLSPADLFSCCEELKOSAKRVEHKFTAAYVHQAFLMQLSHDLLIDILSSDNLYE 189  
146 PNCGLMLLSDAHQCTKLVELSWRMCLSNFQIRKNEDELQLPQDMVVOQLSSSELEFE 205  
190 KEETREAAAMLLENTSRSQYLSVLSQIRIDALSTYQRAWQGLPNDKSVVVOQL 249  
206 DERLYESAMNWSYDLKRRYCYLPQLQTVRLALLPAILYM-----ENVAMEBL 255  
250 YKSMKPFKPRLGMTKE--EMMIFIASSENPCSLYSVCYSPQ-----A 292  
256 ITRKOR-----SKEIYEAIRCKLKLQNDGVVTSICARPKRTGHALFLGGOTFNC 307  
293 EKVY-----KLCSPADL--HKVGTVPDNDIYIAGQVPLKNTKTNHSTKSLQTA 343

Db 308 DKLYVDQAKKEIIPKADIPSPKESFACAIQCKYITGG-----RGSENGVSKD----- 357  
QY 344 FRTVNCFW-FDAQNTWPKPTPMLFVRIKPSLVCCGYIYAGDS-----VGSELN 395  
Db 358 -----VVVYDTLHEBMSKAAAPMLVAFGHSALBKCLYVGGHTAATGCLPASPSVS 410  
QY 396 RRTVERYTEKDEMTWVSPPCAMQMSAAV-----VHDCIYVMTLNLMYCY 442  
Db 411 LKQVEHYDPTINKMTWVAPLREGVSNAAVSAKLKLPAGGTSVSHD-----KLPRKYQCY 465  
QY 443 PRSDSVEMAMQTSRSPASAAAFGDKIFYIGLHIATNSGIRLPSGTVDSSVTEIY 502  
Db 466 DQENRMTVPATCPQPMRYTAAVNGQIFIMG-----GDTEFSACSAYKF 511  
QY 503 DVNKNEMKMAANIIPAKRYSDCPVRVAVVISNLCVFMRETHLNERAKYTYO-----YD 555  
Db 512 NSETYQTKYGDVTAKRMS-----CHAVASGNKLYV-----GGYFGIQCKTLDCYD 559  
QY 556 LELDRMS 562  
Db 560 PTLDVWN 566

RESULT 5  
US-08-862-508-2  
Sequence 2, Application US/08862508  
Patent No. 6066451  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/862,508  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silveri, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEO ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-862-508-2

Query Match 15.5%; Score 511.5; DB 3; Length 589;  
Best Local Similarity 25.4%; Pred. No. 4.3e-44;  
Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

QY 11 TEVAVSLLEBQKLFYQOQLFTDVLVEGTFEPCMKVTLATGSSYRRAMMSGISLSSKQT 70  
Db 26 SSYADSVLTLLNLROORLFTDVLHAGNTPFCBRAVLAACSRFEAFMSGGLKESQDS 85

QY 71 HVHLRN-VDAATLQIITTYATGNLAMDSTVEQLEYTFACFQVEDVLQRCREYLIIKIN 129  
Db 86 EYVFNDSIHPEVLELLELDLDAVYSRVIIYENENESLLENGDMLEFPQDINADACFELEKIH 145  
QY 130 AENCVRLLSFADLFCEBEIKQSAKMEVHEKTPAVYHQDAFMQLSHDLIDILSSONLVE 189  
Db 146 PTNCLGMLLSDAHOCTKLYELSMWCLSNFQTIKRNEDFLQLPQDMVQLLSSELETE 205  
QY 190 KEETVEALMLLENTESRQYLSYSQIRIDLSBVTQAMFQGLPBNKSVYVQGL 249  
Db 206 DERLYESAINMISYDLKRYCYLPBLQTVRLALLPALYLM-----ENVAMEBL 255  
QY 250 YKSMPEFFKRLGWTKE--EMMIFTEASENPSLSYSCYSPQ-----A 292  
Db 256 ITRKOR-----SKEIYEAIRCKLKLQNDGVVTSICARPRTHGALFLGGOTFWC 307  
QY 293 ERYVY-----KLCSPADI--HXYGVTVTPNDIYIAGQVPLKNTKTNHSTKSLQTA 343  
Db 308 DKLYVDQAKKEIIPKADIPSPKESFACAIQCKYITGG-----RGSENGVSKD----- 357  
QY 344 FRTVNCFW-FDAQNTWPKPTPMLFVRIKPSLVCCGYIYAGDS-----VGSELN 395  
Db 358 -----VVVYDTLHEBMSKAAAPMLVAFGHSALBKCLYVGGHTAATGCLPASPSVS 410  
QY 396 RRTVERYTEKDEMTWVSPPCAMQMSAAV-----VHDCIYVMTLNLMYCY 442  
Db 411 LKQVEHYDPTINKMTWVAPLREGVSNAAVSAKLKLPAGGTSVSHD-----KLPRKYQCY 465  
QY 443 PRSDSVEMAMQTSRSPASAAAFGDKIFYIGLHIATNSGIRLPSGTVDSSVTEIY 502  
Db 466 DQENRMTVPATCPQPMRYTAAVNGQIFIMG-----GDTEFSACSAYKF 511  
QY 503 DVNKNEMKMAANIIPAKRYSDCPVRVAVVISNLCVFMRETHLNERAKYTYO-----YD 555  
Db 512 NSETYQTKYGDVTAKRMS-----CHAVASGNKLYV-----GGYFGIQCKTLDCYD 559  
QY 556 LELDRMS 562  
Db 560 PTLDVWN 566

RESULT 6  
PCT-US95-12508-2  
Sequence 2, Application PC/TUS9512508  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/12508  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silveri, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259PC



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-12508-2

Query Match 15.5%; Score 511.5; DB 5; Length 589;  
Best Local Similarity 25.4%; Pred. No. 4.3e-44;  
Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

QY 11 TEVAVLLLEQLKLFYEQQLFTDVLIVLVEGTEPCHKVLTATGSSYFRAMFMSGLSSKOT 70  
DB 26 SSTADSVLTHNLRLRQRLFTDVLHAGNRTFPCRAVLAACSRFEAMFSGGLSKESQDS 85  
QY 71 HHLARN-VDAATLQIITVAVTGNLAMDSTVEQLYETACFLQVEDVLQRCREYLTKIN 129  
DB 86 EVNFDNSIHPEVLELLDVAYSRVLIINENASLEAGDMLFQDIRDCAFLKKNLH 145  
QY 130 AENCVRLLSFADLFSCBELKQSAKRVNFKTAVYHODAFMQLSHDLLIDISSDNLNVE 189  
DB 146 PTNCLGMLLSDAHQCTKLYELSMWCLSNFQIRKNEIDFLQPODMVQQLSSELETE 205  
QY 190 KEETVEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTQRAMFGQLPNDKSVVVOGL 249  
DB 206 DELVATESAINWSTYDKKRYCYLPBELLQTVRLALPAILYM-----ENVAIBEL 255  
QY 250 YKSMPEKPRKLTMTKE--EMMIFIEASSENPCSLYSVCYSPQ-----A 292  
DB 256 ITRKOR-----SKEIYBEAIRCKLKILQNDGVVTSLCARPKTGHALFLGQTFMC 307  
QY 293 EKYV-----KLCSPADL---HKVGVTPDNDIYIAGGVPLKNTKTNHSTSKLQTA 343  
DB 308 DKLYLDOKAKEIIPRADIPSPRKEFSACAIQCKVYITG---RSENGVSD----- 357  
QY 344 FRVNCFFM-FDAQONTWPKTMTLVRIKPSLVCEGTYAIGDS-----VGGELN 395  
DB 358 -----VWYDYLHEBMSGAAPMLVARFGHSAELKFCLYVGGHTAATGCLPASPSVS 410  
QY 396 RRTVERYTEKDEMTWVSPRPCAMOWSAAY-----VHDCIYVMTLNMVYCY 442  
DB 411 LKQVEHYDPTINKTVAAPLRBGSNAVAASAKLQAPRGCTSVSD---KLPRVQCY 465  
QY 443 PPRSDSVEMANRQTSRSPASAAAFDDKIFYIGLHIATNSGIRLPSGTVDSSVTEIY 502  
DB 466 DDCENRWTPATCPQPMRYTAAAVLGNQIFIMG-----GDTEFSACSAVYKF 511  
QY 503 DVNKNEMKMAANIIPARYSDPCRAVVISLGVEMRETHLNRAKYVYQ-----VD 555  
DB 512 NSETTYMTKGVDTAKRMS---CHAAVAGSKLYV-----GGYFGIQRCKTLDCYD 559  
QY 556 LELDRMS 562  
DB 560 PTLDVVN 566

## RESULT 7

US-09-154-750A-86  
Sequence 86, Application US/09154750A  
Patent No. 6432640  
GENERAL INFORMATION:  
APPLICANT: Vogelstein, Bert  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Polyak, Kornelia  
TITLE OF INVENTION: p53-Induced Apoptosis  
FILE REFERENCE: 1107,75357  
CURRENT APPLICATION NUMBER: US/09/154,750A  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/059,153

PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/079817  
PRIOR FILING DATE: 1998-03-30  
NUMBER OF SEQ ID NOS: 93  
SOFTWARE: FastSeq for windows Version 3.0  
SEQ ID NO 86  
LENGTH: 516  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-154-750A-86

Query Match 12.8%; Score 422.5; DB 4; Length 516;  
Best Local Similarity 24.1%; Pred. No. 6.1e-35;  
Matches 151; Conservative 87; Mismatches 220; Indels 169; Gaps 21;

QY 11 TEVAVLLLEQLKLFYEQQLFTDVLIVLVEGTEPCHKVLTATGSSYFRAMFMSGLSSKOT 70  
DB 26 SSTADSVLTHNLRLRQRLFTDVLHAGNRTFPCRAVLAACSRFEAMFSGGLSKESQDS 85  
QY 71 HHLARN-VDAATLQIITVAVTGNLAMDSTVEQLYETAC-----FLQVEDVLQ 119  
DB 86 EVNFDNSIHPEVLELLDVAYSRV-----IHQL-EGKCRSLGSLVTCSPFDIIRA 138  
QY 120 CREYLTKINAEVRLSFADLFSCBELKQSAKRVNFKTAVYHODAFMQLSHDLLID 179  
DB 139 CAEFLKKNLHPNCLGMLLSDAHQCTKLYELSMWCLSNFQIRKNEIDFLQPODMVQ 198  
QY 180 IISDNLNVEKETVEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTQRAMFGQLP 239  
DB 199 LISSELETERLVESAINWSTYDKKRYCYLPBELLQTVRLALPAILYM----- 250  
QY 240 NDKSVVVOGLYKSMPEKPRKLTMTKE--EMMIFIEASSENPCSLYSVCYSPQAKVYK 297  
DB 251 --ENVAIBELITRKOR-----SKEIYBEAIRCKLKILQNDGVVTSLCARP----- 294  
QY 298 LQSPPADLHKVGVTPDNDIYIAGGVPLKNTKTNHSTSKLQTAFTVNC--PYWEDA 355  
DB 295 -----KTG-----HALFLGQ-----TFMCDKLYLVQD 318  
QY 356 QONTWPKTMTLVRIKPSLV---CEGYIYAGDSVGGELNRTVER---YDTEKDE 408  
DB 319 KKEIIPKADISPR-KEFSACAIQCKVYI-----TGRGSENGVSDVWYDYLHEE 370  
QY 409 KTWVSPRPCAMOWSAAYVVHDCIYVMTLNMVYCYPRSDSVEMANRQTSRSPASAAAG 468  
DB 371 WSKAPMLVARFGHSAELKCLYV----- 395  
QY 469 DKIFYIGLHIATNSGIRLPSGTVDSSVTEIYDVNKNEMKMAANIIPARYSDPCRAV 528  
DB 396 -----VGGHTAATGCLPASPSVSLK---QVEHYDPTINKTVAAPRRRRYV--CAQV 444  
QY 529 VTSNLSVEMRETHLNRAKYVYQYDDELDMWSLRHISERVNLVDGDFCTYCKLYP 588  
DB 445 SAKLKLPAFGGTSVSHDKLPKQV--YDCCENRW-----TVPATCP 483  
QY 589 SCLSESPMKPPTYLPSTDGTETFEIDG 615  
DB 484 -----QPMRIHQASCPSGTQDPFLMG 505

## RESULT 8

US-09-337-307A-3  
Sequence 3, Application US/09337307A  
Patent No. 6432692  
GENERAL INFORMATION:  
APPLICANT: Bradfield, Christopher A.  
APPLICANT: Carver, Lucy A.  
APPLICANT: Dunham, Elizabeth E.  
TITLE OF INVENTION: Sensitive Biosay for Detecting Agnists of The Aryl Hydrocarbon  
FILE REFERENCE: MAR0105  
CURRENT APPLICATION NUMBER: US/09/337,307A  
PRIOR FILING DATE: 1999-06-21  
NUMBER OF SEQ ID NOS: 19



```

1 FILE REFERENCE: 026579-186
2 CURRENT APPLICATION NUMBER: US/09/068,655A
3 CURRENT FILING DATE: 1998-08-31
4 EARLIER APPLICATION NUMBER: AU PN 6587
5 EARLIER FILING DATE: 1995-11-15
6 EARLIER APPLICATION NUMBER: PC/AU96/00725
7 EARLIER FILING DATE: 1996-11-15
8 NUMBER OF SEQ ID NOS: 13
9 SOFTWARE: PatentIn Ver. 2.0
10 SEQ ID NO 6
11 LENGTH: 553
12 TYPE: PR1
13 ORGANISM: Myxoma virus
14
15 US-09-068-655-6

```

Query Match 10.0%; Score 329; DB 4; Length 553;  
Best Local Similarity 21.2%; Pred. No. 3.6e-25;  
Matches 131; Conservative 85; Mismatches 231; Indels 172; Gaps 21;

QY 13 YASLLLEQLLFEQOOLFEDVILIV-EGTEPCHKKVLAQCSYPRAFMGLSSKQTH 71  
 Db 5 YAIRLLESTRINLODKTKLCDVLTLYTDDVDSIAHAKILLISASTYETEFMFSDIFEKORNV 64  
 QY 72 VHLRNDAATLQIITTYAYTGNLAMDSTVEQLYETACP,QVEDVLOCREYLKIKNAE 131  
 Db 65 INV-CVEYRALHLINFIYSGLTRLTDDPDVDCIIVAADVLQILSELSAENFILARLAE 122  
 QY 132 NCYRLLSFPADLFSCEELKOSAKRMVEHKTATVYHODAFMQSLHDLLIDILSDNINYEKE 191  
 Db 124 NCLHYEYFSKRYNRHRIFFNVYITTLIHNIIVSVLROPNFEKTELCOLQNIILSSDDLNVIDE 183  
 QY 192 ETYREAAMLLEVTNTRSQOYLSSVLSQIRIDALSE-----VTQAMPQGL--- 23  
 Db 184 DVCAVVLVTMLKON--NMEDCPSEVLEQVRMSLLSMYKILLKTPCIRNRKYVQSLAKL 241  
 QY 238 -----PNDKSV----- 244  
 Db 242 DHSRPRPTQCLISIGGRKYDDITSSPELYKSPDDVWTYVSYLPTHRQEPSVAVLDFV 301  
 QY 245 --VVGGL-----YKSNPKFFKRLG--MTKEEMIFIEASSENPCS 281  
 Db 302 VVVVGGLODSVASVSVYDKTNEMKECPPLKSPRHGGVLVLADKLIVIGKGRN--S 355  
 QY 282 LYSVVCY-SPOAEKRYKLCSPPADLHKGTVTVPNDIYVAGGQVPLKNTYTHNSKTSKL 344  
 Db 360 YLNDVYMPETAYATWRKCLSLREARTNVAALV--RNKYTYTIGG--IRSEV---EPSRL 411  
 QY 341 QTAFTFVNCYMWEDAQONTMFPK-----TEMLFVRKPSLVCEEGYVIAIGDSVGBELNR 396  
 Db 412 E-CITVECL-----ONNKNVAAKKSIPBEKACLAAP---YKHETVYAGGYALNG--R 456  
 QY 397 RTVERDYDEKDEWMTWSPLPCAMQMSAAVVHDCIYVMTLNLNMYCYPFRSDSWEMAMRO 456  
 Db 459 GTV-----VTKENTLYMIVNELDDMTYLLPME 485  
 QY 457 TSSRSPASAAFGDKIFYIGGLHATNIGRLPSGTVDSSVTELYDVNKNQEMKMAANIP 516  
 Db 486 LSHNDASLGVLGKDLVYVGGF-----VGSQGTNSVEKYNKNTNNBRILIPCK 533  
 QY 517 AKRYSDPCRAVAVISNLC 535  
 Db 533 SPKYG---HCSVVLNHEC 547

RESULT 11  
 US-08-307-499-29  
 Sequence 29, Application US/08307499  
 Patent No. 5651972  
 GENERAL INFORMATION:  
 APPLICANT: Moyet, Richard W.  
 APPLICANT: VJ uela, Eladio  
 APPLICANT: Gibbs, E.P.J.  
 TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a

TITLE OF INVENTION: Live Vaccine Vector  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

Query Match	9.4%	Score 309.5	DB 1	Length 530
Best Local Similarity	21.7%	Pred. No. 3.5e-23		
Matches 125	Conservative 98	Mismatches 215	Indels 139	Gap 22
Qy	16	SLLEQLKLFYRQGLFTD--LYLYEGTEFFPGHKWVLACSSFFRAMPFSGLSSEKQTNH	73	
Db	3	SLLRRLDFFHGIGMCDIKIYSIENKNTISARHLISMYSKFYFNIIPSDFDKNDDEY	62	
Qy	74	LRNVDAAFTQLIIIIYAATGNLANMDSVTEQLYETACFQVEDVQCRGREYLKIKINAENC	133	
Db	63	I-CADYDILYLIIFEMTGNITVLTKDNIEVLQVCDYLCIDSLIKICEBYIGIIDETNC	121	
Qy	134	VRLSPADLFSCBELKOSAK-----RMV-----EHKFTA-----	162	
Db	122	IHLNFSDTVLQRLREMSKWLPEKLIINNKLVELDIDMLIKIYKTIACBYIVKKI	181	
Qy	163	----VYHODAFMQLSHDLILDLISDNL-----NKEKETVREAAMLMEYNTSRQYLS	214	
Db	182	ILNNIVKHDEIITYTKLMLKHIINDOHTLSLDIELYNNIBRIYDKHEDVDISHNFI	241	
Qy	215	SVLSQ--IRIDALSEVTQ-----RAWFQGLPPNDKSVVVGLYKSNPKPEK	258	
Db	242	MVGSKKILFNITAFNPLSNKKHIIIDRYDMFPCCKTHFSVYVLSIYIYIGC-----K	292	
Qy	259	PRLGMTKEEMNIFIEASSENPCSLYSVCSYSPQAEKVKYKCSPPADLHKVGTVPDPND	317	
Db	293	KRGVFTKEVLSYNKKN-----KLM-----CYEPE-----LNFYRDIYSVCVSNMG	332	
Qy	318	IYIAGGVPLKNYITNHSKTSKLOTAFRTVNCYFWDAQONTWPKPTDMLFVRIKPSLVC	377	

Db 333 IYISIGK-----DTNGYMTNIVE-----FM-KPEKWSYDGOHLCPRCYMSLVD 376  
Qy 378 CEGYIYAIG--DSVGEINRRTVERD--TEK---DEWTNVSPLPCAMQMSAAVYVHDCI 431  
Db 377 YNNEVYTIIGLKTSTIDEFNIEIMVSDDAVEKLTDSMMKLPKPLAKSGISSIYVNDFI 436  
Qy 432 Y-----VMTLNLMYCFPRSDSVEMAMRQTSRSPASAAAFGDKIFYIGLHI 479  
Db 437 YCIGRIDTPHISIEHTNDVYIYSSRDCKMKYSTNNVARSFCLSCVFNNELIYIGYN- 495  
Qy 480 ATNSGIRLPSGTVDSVTVETVDNKNEMKMAANIP 516  
Db 496 -TNS-----VEKYNKLTWTKRLNDIP 516

RESULT 12  
US-09-299-268-29  
Sequence 29, Application US/09299268  
Patent No. 6217882  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Gibbs, E. P. J.  
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/299,268  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/901,127  
FILING DATE:  
FILING DATE: 1-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,630  
FILING DATE: 29-JUN-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/342,212  
FILING DATE: 21-APR-1992  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF35.1.FWCCI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 530 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-299-268-29

Query Match 9.4%; Score 309.5; DB 4; Length 530;  
Best Local Similarity 21.7%; Pred. No. 3,5e-23;

Matches 125; Conservative 98; Mismatches 215; Indels 139; Gaps 22;  
Qy 16 SLLEQLKFEYEQQLFTD--IYLVIGTEPPCHKMLATCSSFARMFSGSSEKQTHVH 73  
Db 3 SLLERLHDFKKGICWDIKIYISINNKTISARHLTILMSYSKYFYNI FNSDFIDKNNDELY 62  
Qy 74 LNNVDAATLQIITTYAIGNLAMDSTVEQLYETACFLQVEDVYLQRCREYLKIKINAENC 133  
Db 63 I-CADYDILYIILFEMTYGNIVLTQNDIELVQVCDYLCIDBLINICEBYIGIILDENIC 121  
Qy 134 VRLSPADLFSCBELKQSAK-----RMV-----EHKFTA----- 162  
Db 122 IHLNFSPTDYNLQRLREMSKWYLPKIINNKLIVLELIDIMILLIKEIKIYACEYIVKKI 181  
Qy 163 ----VYHODAFMQSLHDLLIDILSSDNL-----NVEKETVREAAMLMEYNTESRSQYLS 214  
Db 182 ILNWIYHDERIITYKKLKHINDODHTLSLSDIELYNNIRIRIYDNKEHDVISHNFI 241  
Qy 215 SVLSQ--IRIDALSEVTO-----RAWFQGLPPNDKSVVVOGLYKSNPKPEK 258  
Db 242 MWGKKKIFNITAFNPLSNKKHIIIDRYDMFGCKTHFSVYVLSIYIIG-----K 292  
Qy 259 PRLGNTKERMMIFIASSENPCSLYSVCYSPQAEKRYKLCSPPADLHKVGTVPDND- 317  
Db 293 KRGYFTKEVLSYNIN-----KLV-----CYEBE-----INRYRYDTSVCVSNQM 332  
Qy 318 IYIAGGOVPLKNTKTNHSTSKLOTAPFTVNGFYVPAQOQNTWPPKTMLPVIRKPSLYC 377  
Db 333 IYISIGK-----DTNGYMTNIVE-----FM-KPEKWSYDGOHLCPRCYMSLVD 376  
Qy 378 CEGYIYAIG--DSVGEINRRTVERD--TEK---DEWTNVSPLPCAMQMSAAVYVHDCI 431  
Db 377 YNNEVYTIIGLKTSTIDEFNIEIMVSDDAVEKLTDSMMKLPKPLAKSGISSIYVNDFI 436  
Qy 432 Y-----VMTLNLMYCFPRSDSVEMAMRQTSRSPASAAAFGDKIFYIGLHI 479  
Db 437 YCIGRIDTPHISIEHTNDVYIYSSRDCKMKYSTNNVARSFCLSCVFNNELIYIGYN- 495  
Qy 480 ATNSGIRLPSGTVDSVTVETVDNKNEMKMAANIP 516  
Db 496 -TNS-----VEKYNKLTWTKRLNDIP 516

RESULT 13  
US-08-307-499-20  
Sequence 20, Application US/08307499  
Patent No. 5651972  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Vi uela, Eladio  
APPLICANT: Gibbs, E. P. J.  
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/307,499  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,241  
FILING DATE: 1-JUL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,630  
FILING DATE: 29-JUN-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/342,212  
FILING DATE: 21-APR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF35.1.FWCC1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-307-499-20

Query Match 7.1%; Score 234.5; DB 1; Length 500;  
Best Local Similarity 20.4%; Pred. No. 2e-15;  
Matches 120; Conservative 92; Mismatches 222; Indels 155; Gaps 25;

1 MSTDQROINTEYAVSL-LEQLKLFYEQOLFDTIVLIVEGTEPCHKWLATCSSYFRAM 59  
1 MSKQ-ETIDYVYIERLAVNLRSDYDEI--VFIMVGAVVKKKELVSVSNYFKLI 56  
60 FMSGSESQKQTHVLRNVDATLQIITAYTGNLAMDSTYEQLYSFACFLQVEDVLR 119  
57 TKNQSNELTVSPQSFYFLD-----IKYIEGTIVTIDLNVENIFSISSKADPLKNS 110  
120 CREYLIKKNAENCRLLSPADLFSCEELKQSKAKWEHKTAVYHQDAFMQLS-HDLLI 178  
111 CIDPMSKAITDSTCVKIKYIKGFNGCFAYVNDIAVIRKFTKI-ETDILSLSLPDLRI 169  
179 DILSSDNINVEKETVRBAAMLMEYNTESRQYLSVLSQIRIDALS-----EYTORAM 233  
170 -ILKSGELDVSEDDVLLPFIKMSRHKKSNNRKSFTLVETVLRKYNLSTYGYKLT-K-W 226  
234 FQGLPNDKSVVVOGLYKSNP-----KFFKPRLCM-----TKEMMIFI 272  
227 LARFGKNNN---VELNENELPRISYQHRFTNRKTYMTVPESPSINMLGNVSVNLSIIN 283  
273 E-ASSENP-CS-LYSSVCY-----SPOAEKYKLCSPPADLH--KVG 310  
284 SIAENHNPCYCGSVLAMDILYLIGINKSLDPVSDITSVDTRSFIEHTPPL-LHPRKCPG 342  
311 VTPDNDIYIAGS---QVPLKNTKTHSKTSKQTAFRITNCTCYWDAQNTMFPKTPML 367  
343 VALFKRRIYVGGIGYDGPLKTVES-----WSPGEQO--RREVP 382  
368 FVRIKPSLVCESEGYIAYIGDSVGLNRRTERVRYTEKLEMTVMVPLPCAMQMSAAVVV 427  
383 QRFNFCIIGTDNDLVVGGISF---DKTIEIYEERTWISIGNAMYSHGCGIAIH 438  
428 HDCIYVMTLNLAKCYPRSDSVWEMARQTSRSPASAAFGDKI FYIGGLHATNSGIRL 487  
439 HGIIYV-----IGGLSFDNIHV-- 456  
488 PGGTVGSSVTVETVDNKNEMGMANIPAKRYSDCVRAVVISNGLCV 536  
457 -----FTWVEKINPHSKMTVEKSLPPRF-----NSSLCI 487

RESULT 14  
US-09-299-268-20  
Sequence 20, Application US/09299268  
Patent No. 6217882  
GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.  
APPLICANT: Vi uela, Eladio  
APPLICANT: Gibbs, E.P.J.  
TITLE OR INVENTION: Use of Recombinant Swine Foxvirus as a  
TITLE OF INVENTION: Live Vaccine Vector  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/299,268  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/901,127  
FILING DATE:  
APPLICATION NUMBER: US 07/908,241  
FILING DATE: 1-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,630  
FILING DATE: 29-JUN-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/342,212  
FILING DATE: 21-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF35.1.FWCC1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-299-268-20

Query Match 7.1%; Score 234.5; DB 4; Length 500;  
Best Local Similarity 20.4%; Pred. No. 2e-15;  
Matches 120; Conservative 92; Mismatches 222; Indels 155; Gaps 25;

1 MSTDQROINTEYAVSL-LEQLKLFYEQOLFDTIVLIVEGTEPCHKWLATCSSYFRAM 59  
1 MSKQ-ETIDYVYIERLAVNLRSDYDEI--VFIMVGAVVKKKELVSVSNYFKLI 56  
60 FMSGSESQKQTHVLRNVDATLQIITAYTGNLAMDSTYEQLYSFACFLQVEDVLR 119  
57 TKNQSNELTVSPQSFYFLD-----IKYIEGTIVTIDLNVENIFSISSKADPLKNS 110  
120 CREYLIKKNAENCRLLSPADLFSCEELKQSKAKWEHKTAVYHQDAFMQLS-HDLLI 178  
111 CIDPMSKAITDSTCVKIKYIKGFNGCFAYVNDIAVIRKFTKI-ETDILSLSLPDLRI 169  
179 DILSSDNINVEKETVRBAAMLMEYNTESRQYLSVLSQIRIDALS-----EYTORAM 233  
170 -ILKSGELDVSEDDVLLPFIKMSRHKKSNNRKSFTLVETVLRKYNLSTYGYKLT-K-W 226  
234 FQGLPNDKSVVVOGLYKSNP-----KFFKPRLCM-----TKEMMIFI 272  
227 LARFGKNNN---VELNENELPRISYQHRFTNRKTYMTVPSSPSINMLGNVSVNLSIIN 283

```
QY 273 E-ASSENPC-CS--LYSSVCY-----SPOAEKRYKLCSPADLIH--KVGT 310
      |||
Db 284 SIAMENHPVCGSVLMDILVIGINKSLDPVDISVDRSRFELHTPL-LHPKCPG 342
      |||
QY 311 VTPDNDIYAGC---OVLKNTKTNHSTKSLQTAFTYNCYWFPAQONTFPPTML 367
      |||
Db 343 VAFKRIYVVGIGYDGPLKTVES-----WSPGEQO-WREEVPLL 382
      |||
QY 368 FVRIKPSLVCCEGIYVIGDSVSGELNRRTRVRYDTEKDEMTVSPPCAMQMSAAVY 427
      |||
Db 383 QPFPNCIITDNDLYVVGISE---DDKTIETYEENTWSIGNAMNYSHPGCCIAH 438
      |||
QY 428 HDCIYVWTLNLMYCYPFRSDSWVEMAMRQTSRSPASAAFGDKI FYIGLIHATNSGIRL 487
      |||
Db 439 HGIIYM-----IGLSFIDNIHV-- 456
      |||
QY 488 PSCGTVDGSSVTVEIYDVNKNEMKMANIPAKRYSDCPCRAVVISNGLCV 536
      |||
Db 457 -----FTMVEKYNPHSNKMTVEKSLPFRF-----NSSLCI 487
      |||
```

## RESULT 15

```
US-09-137-223A-2
/ Sequence 2, Application US/09137223A
/ Patent No. 6420525
/ GENERAL INFORMATION:
/ APPLICANT: Yee, David P
/ APPLICANT: Delisher, Theresa A
/ TITLE OF INVENTION: TESTIS-SPECIFIC TRANSCRIPTION FACTOR
/ FILE REFERENCE: ZGCL-1
/ CURRENT APPLICATION NUMBER: US/09/137,223A
/ CURRENT FILING DATE: 1998-08-19
/ PRIOR APPLICATION NUMBER: 06/056,130
/ PRIOR FILING DATE: 1997-08-19
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 478
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-09-137-223A-2
```

Query Match 5.9%; Score 195.5; DB 4; Length 478;  
Best Local Similarity 21.9%; Pred. No. 2,1e-11;  
Matches 84; Conservative 58; Mismatches 149; Indels 93; Gaps 13;

```
QY 4 QDERQJNTYAVSLLEQLKLFYEQQLF---TDIVLIVGTEPCHKNVLTATCSSYFRA 58
      |||
Db 41 EQORLNTTPRRKRLKTSKYIY-QTLFLNGENSDIKICALGHEWSLHKIYLCO-SGYFS 98
      |||
QY 59 MEMSGLSESKOTVHL-----RNVDAATLOIITVATGNLANMDSVTEQLYETACFLQVE 114
      |||
Db 99 MFGSGWKESSMNTIELEIPDONTDEALQVAFGSLYRDVDLIKPSRVVALAALAAQLQUD 158
      |||
QY 115 DVLQRCREYLIIKINAENCVRLLSPADLFSCBELKQSAKRMVSHKFTAVYHODAFMQLSH 174
      |||
Db 159 GLIQCGGEMTKETVNVVTCGYTSACTYGLDSVKKCKLEWLNLMTHQNVLELFKELSI 218
      |||
QY 175 DLLIDILSDNLNV-EKETVRBAAMLWL-----BYNTES----- 208
      |||
Db 219 NVMKQLIGSSNLFVMQVEMDIYALKKMPLQLVPSWNGSLKQLLTETDVMFSQRRKPE 278
      |||
QY 209 -----RSQYLSVLSQIRI--DA-----LSEVTQRAWFOGL-P 238
      |||
Db 279 GMAFLETEQCKPVSFRRHLRLQIITISDLASARIIEODAVPSEWLSVYKQWFMALRA 338
      |||
QY 239 PNDKSYVVOGLYKSMPEFFKPRLG--MTXB-----EMTIFIEA 274
      |||
Db 339 EODSEVGPQEIINKELEBEGNSMRGCKRLADGEYCWMTGTFNPGFDLLVYTRNIIIFKRN 398
      |||
QY 275 SSENPCSLYSSVCYSPQAEKRYKL 298
```

Db 399 TLNOPS--GSVLOPFRRSIAFRL 420  
Search completed: July 14, 2003, 18:24:28  
Job time : 9.59422 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 / Search time 21.5998 Seconds  
(without alignments)  
3843.329 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292

Sequence: 1 MSTDQERQINTEYAVSLLEQ.....STDGHEPFLDGMVALPVV 623

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seque, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3292	100.0	623	22	AAE11886
2	3292	100.0	623	22	AAE11886
3	3204.5	97.3	622	23	AAE60081
4	1675	50.9	335	21	AAE42737
5	578.5	17.6	623	22	AAE62185
6	569.5	17.3	816	22	ABG03507
7	558.5	17.0	609	22	AAE92953
8	555.5	16.9	689	22	ABE63086
9	554	16.8	606	23	ABE90399
10	541.5	16.4	587	22	AAE38956

11	534	16.2	593	22	AAE78511
12	530	16.1	749	22	AAE79495
13	526.5	16.0	754	22	ABE60307
14	523	15.9	575	23	AAE47944
15	523	15.9	748	22	AAE40225
16	516.5	15.7	589	17	AAE94389
17	512.5	15.6	596	23	ABE60123
18	511.5	15.5	589	17	AAE94386
19	509	15.5	532	21	AAE94936
20	508.5	15.4	584	22	AAE41870
21	505.5	15.4	569	22	ABE11665
22	505.5	15.4	569	22	AAE40204
23	505.5	15.4	569	22	AAE41990
24	505.5	15.4	569	22	AAE94214
25	505.5	15.4	569	23	AAE94302
26	501.5	15.2	521	22	ABE76659
27	501.5	15.2	569	23	AAE79005
28	501	15.2	626	22	AAE97108
29	500	15.2	568	22	AAE94018
30	500	15.2	568	22	AAE94486
31	498	15.1	625	22	ABE59896
32	484.5	14.7	574	22	ABE64408
33	484	14.7	574	22	AAE87111
34	480.5	14.6	575	22	AAE71239
35	478.5	14.5	610	22	AAE40497
36	474	14.4	617	23	AAE97212
37	470	14.3	565	23	ABE81893
38	462	14.0	597	22	AAE93687
39	460.5	14.0	538	22	ABE67078
40	457	13.9	604	22	AAE10610
41	455.5	13.8	584	22	AAE95583
42	455	13.8	538	22	AAE16297
43	453	13.8	506	21	AAE43090
44	435.5	13.3	616	22	AAE95123
45	435.5	13.2	615	22	AAE28187

## ALIGNMENTS

RESULT 1	AAE11886	standard; Protein; 623 AA.
ID	AAE11886	
AC	AAE11886	
DT	18-DEC-2001	(first entry)
DE	Angiogenesis associated kelch-like protein (KLP).	
KW	Angiogenesis associated protein; AAF; cytosolic; cardiac; gene therapy; ophthalmological; vulvar; myocardial infarction; macular degeneration; diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine; rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation; cancer; therapeutic; diagnostic; kelch-like protein; KLP.	
OS	Unidentified.	
PN	WO200170808-A2.	
PD	27-SEP-2001.	
PF	22-MAR-2001; 2001WO-US09609.	
PR	22-MAR-2000; 2000US-191134P.	
PA	(CURA-) CUREGEN CORP.	
XX	(GETH) GENENTECH INC.	
PI	Rastelli LK, Gerritsen M;	
DR	WPI, 2001-602775/68.	
DR	N-PSDB; AAD19115.	

XX Novel angiogenesis associated polypeptides and polynucleotides encoding  
 PT the polypeptides, useful for modulating angiogenesis and for treating  
 XX tumors and cancers -  
 PS Claim 1, Page 9-11, 159pp, English.  
 XX  
 CC The invention relates to angiogenesis associated proteins (AAP) and their  
 CC corresponding cDNA molecules, which are useful for modulating  
 CC angiogenesis. AAP proteins and nucleic acids are useful for promoting  
 CC wound healing, for example after organ transplantation, and in the  
 CC treatment of tumours, myocardial infarction, cancers, diabetic  
 CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
 CC AAP proteins and DNA's are useful in potential prophylactic and  
 CC therapeutic applications implicated in a variety of disorders including  
 CC those related to angiogenesis, and also in diagnostic applications.  
 CC AAP cDNA is also useful in gene therapy. The invention also relates to  
 CC a method for screening a tissue sample for tumorigenic potential. AAP  
 CC proteins are used to screen drugs or compounds that modulate AAP activity  
 CC or expression as well as treating disorders characterised by insufficient  
 CC or excessive production of AAP or production of AAP forms that have  
 CC decreased or aberrant activity compared to the wild type protein, or  
 CC modulate biological function that involve AAP. The present sequence is  
 CC Kech-1 like protein (KLP) which is an angiogenesis associated protein  
 CC (AAP) of the invention. KLP is associated with tube formation and  
 CC angiogenesis because it is upregulated in the in vitro model of  
 CC angiogenesis. Kech mediates cytoskeletal associations, it is involved in  
 CC morphogenic processes such as tube formation, that depend on  
 CC cytoskeletal arrangements and signalling.  
 XX  
 SQ Sequence 623 AA;  
 Query Match 100.0%; Score 3292; DB 22; Length 623;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;  
 Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSTQDERQINTYEAVALSLLEQLKLFYEQQLFTDVLIVGTEFPCHKMVLATCSSYFRAMF 60  
 DB 1 MSTQDERQINTYEAVALSLLEQLKLFYEQQLFTDVLIVGTEFPCHKMVLATCSSYFRAMF 60  
 QY 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120  
 DB 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120  
 QY 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 DB 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 QY 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 DB 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 QY 181 LSSDNLNVEKEETVREAAMLMLEYNTESRSQYLSVLQIRIDALSEVTQRANFQGLPPN 240  
 DB 181 LSSDNLNVEKEETVREAAMLMLEYNTESRSQYLSVLQIRIDALSEVTQRANFQGLPPN 240  
 QY 241 DKSIVVVOGLYKSNPKPKPRLGNTKEEMAFIASSENPCLSYSCVCSQAQKVKYKLS 300  
 DB 241 DKSIVVVOGLYKSNPKPKPRLGNTKEEMAFIASSENPCLSYSCVCSQAQKVKYKLS 300  
 QY 241 DKSIVVVOGLYKSNPKPKPRLGNTKEEMAFIASSENPCLSYSCVCSQAQKVKYKLS 300  
 DB 241 DKSIVVVOGLYKSNPKPKPRLGNTKEEMAFIASSENPCLSYSCVCSQAQKVKYKLS 300  
 QY 301 PPADLHKVGVVTPPNDIYIAGGOVPLKTKTNTNHSKTSKLQTAFFRTNCGYWFDAQONTM 360  
 DB 301 PPADLHKVGVVTPPNDIYIAGGOVPLKTKTNTNHSKTSKLQTAFFRTNCGYWFDAQONTM 360  
 QY 301 PPADLHKVGVVTPPNDIYIAGGOVPLKTKTNTNHSKTSKLQTAFFRTNCGYWFDAQONTM 360  
 DB 301 PPADLHKVGVVTPPNDIYIAGGOVPLKTKTNTNHSKTSKLQTAFFRTNCGYWFDAQONTM 360  
 QY 361 FPKTPMLFVRKIPKSLVCCGEYIYALIGDSVGEGLNRRTERVYDTEKDEMTWSPPLCCAMQ 420  
 DB 361 FPKTPMLFVRKIPKSLVCCGEYIYALIGDSVGEGLNRRTERVYDTEKDEMTWSPPLCCAMQ 420  
 QY 421 WSAUVVHDCITYMTLNTLMCTFPPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480  
 DB 421 WSAUVVHDCITYMTLNTLMCTFPPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480  
 QY 421 WSAUVVHDCITYMTLNTLMCTFPPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480  
 DB 421 WSAUVVHDCITYMTLNTLMCTFPPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480  
 QY 481 TNSGIRLPSTGVGSSVTVETIYDVNKNEMMANIIPAKRYSDDPCVAVVISNLCVFMKE 540  
 DB 481 TNSGIRLPSTGVGSSVTVETIYDVNKNEMMANIIPAKRYSDDPCVAVVISNLCVFMKE 540  
 QY 541 THLNERAKVTVYQYDLEDRMSLRQHSERVLMDLGRDPRCTVQKLYPSCLEESPMKPT 600

DB 541 THLNERAKVTVYQYDLEDRMSLRQHSERVLMDLGRDPRCTVQKLYPSCLEESPMKPT 600  
 QY 601 YLPSTDGTEEFELDGMVALPV 623  
 DB 601 YLPSTDGTEEFELDGMVALPV 623  
 RESULT 2  
 AAB60081  
 ID AAB60081 standard; Protein; 623 AA.  
 XX  
 XX AAB60081;  
 AC  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE Human transport protein TPTP-1.  
 XX  
 XX Human; transport protein; TPTP; transport disorder; metabolic disorder;  
 KW neurological disorder; cardiovascular disorder; reproductive disorder;  
 KW immune disorder; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200078953-A2.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 16-JUN-2000; 2000MO-US16668.  
 XX  
 PR 17-JUN-1999; 99US-0139923.  
 PR 10-AUG-1999; 99US-0148177.  
 PR 18-AUG-1999; 99US-0149357.  
 PR 28-OCT-1999; 99US-0162287.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Lal P, Yang H, Hallman JL, Tang YT, Bandman O, Burford N;  
 PI Baughn MR, Ajimzai Y, Lu DM, Au-Young J, Patterson C;  
 DR N-PSDB; AAF27701.  
 DR WPI; 2001-041424/05.  
 XX  
 PT Isolated polypeptide with a human transport protein sequence is useful  
 PT for the diagnosis, prevention and treatment of disorders associated  
 PT with the immune, reproductive and cardiovascular systems -  
 XX  
 PS Claim 2; Page 106-107; 165pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for 43  
 CC novel human transport proteins (designated TPTPs). These can be used in  
 CC the diagnosis and treatment of transport, metabolic, neurological,  
 CC reproductive, cardiovascular and immune disorders, and cell proliferative  
 CC disorders such as cancer.  
 CC  
 SQ Sequence 623 AA;  
 Query Match 100.0%; Score 3292; DB 22; Length 623;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;  
 Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSTQDERQINTYEAVALSLLEQLKLFYEQQLFTDVLIVGTEFPCHKMVLATCSSYFRAMF 60  
 DB 1 MSTQDERQINTYEAVALSLLEQLKLFYEQQLFTDVLIVGTEFPCHKMVLATCSSYFRAMF 60  
 QY 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120  
 DB 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120  
 QY 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 DB 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 QY 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180  
 DB 121 REYLKIKINAENCVRLLSPADLFSCEELKQSAKRWVEHKFTAVYHODAFMQLSHDLLIDI 180



```

QY 181 LSSDNLNVEKETVEEAAMLWLENTSSQYLSVLSQIRIDALSEVTORAMFQGLPEN 240
DB 181 LSSDNLNVEKETVEEAAMLWLENTSSQYLSVLSQIRIDALSEVTORAMFQGLPEN 240
QY 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
DB 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
QY 301 PPADLHKVGTVTVPNDIYIAGQVPLKNTKTNHSTSKLQTFRTVNCFFYFMDAQONTM 360
DB 301 PPADLHKVGTVTVPNDIYIAGQVPLKNTKTNHSTSKLQTFRTVNCFFYFMDAQONTM 360
QY 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPPLCAMQ 420
DB 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPPLCAMQ 420
QY 421 NSAAVVVHDCIYVMTLNTMICYFPKPSDSWVEAMKQTSRSPASAAAFGDKIFYIGLHIA 480
DB 421 NSAAVVVHDCIYVMTLNTMICYFPKPSDSWVEAMKQTSRSPASAAAFGDKIFYIGLHIA 480
QY 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAAVVISNLCVPMKE 540
DB 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAAVVISNLCVPMKE 540
QY 541 THLNERAKYVTVYQYDLELDRMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
DB 541 THLNERAKYVTVYQYDLELDRMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
QY 601 YLFSTDTGTEEFELDGEVVALPPV 623
DB 601 YLFSTDTGTEEFELDGEVVALPPV 623

```

## RESULT 3

AA049413  
ID AA049413 standard; Protein; 622 AA.

AA049413;  
08-OCT-2002 (first entry)

Human mammary gland susceptible gene-encoded protein 68.42.

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

Human mammary gland susceptible gene-encoded protein 68.42;

CC the protein, and the use of the protein, gene and antagonist in  
CC therapeutic applications. Mammary gland susceptible gene-encoded protein  
CC 68.42 can be used in the treatment of a variety of diseases such as  
CC embryonic development disorders and tumours. The present sequence  
CC represents human mammary gland susceptible gene-encoded protein 68.42.

XX SQ Sequence 622 AA;

Query Match 97.3%; Score 3204.5; DB 23; Length 622;  
Best Local Similarity 97.9%; Pred. No. 2e-289;  
Matches 610; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

```

QY 1 MSTODERQINTEAVASLLEQLKLFYEQQLFTDILVIVGTEBPCHKAVLATSSYFRAMF 60
DB 1 MSTODERQINTEAVASLLEQLKLFYEQQLFTDILVIVGTEBPCHKAVLATSSYFRAMF 60
QY 61 MSGLESKQTHVHLNNDAAITQIIITTAATGNLAMDSTVEQLYETKACFLOVEDVLQRC 120
DB 61 MSGLESKQTHVHLNNDAAITQIIITTAATGNLAMDSTVEQLYETKACFLOVEDVLQRC 120
QY 121 REYLKKINAENCVRLLSPADLFSCBELKQSAKRVHEKFTAVYHQDAFMQSHDLIDI 180
DB 121 REYLKKINAENCVRLLSPADLFSCBELKQSAKRVHEKFTAVYHQDAFMQSHDLIDI 180
QY 181 LSSDNLNVEKETVEEAAMLWLENTSSQYLSVLSQIRIDALSEVTORAMFQGLPEN 240
DB 181 LSSDNLNVEKETVEEAAMLWLENTSSQYLSVLSQIRIDALSEVTORAMFQGLPEN 240
QY 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
DB 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
QY 301 PPADLHKVGTVTVPNDIYIAGQVPLKNTKTNHSTSKLQTFRTVNCFFYFMDAQONTM 360
DB 301 PPADLHKVGTVTVPNDIYIAGQVPLKNTKTNHSTSKLQTFRTVNCFFYFMDAQONTM 360
QY 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPPLCAMQ 420
DB 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPPLCAMQ 420
QY 421 NSAAVVVHDCIYVMTLNTMICYFPKPSDSWVEAMKQTSRSPASAAAFGDKIFYIGLHIA 480
DB 421 NSAAVVVHDCIYVMTLNTMICYFPKPSDSWVEAMKQTSRSPASAAAFGDKIFYIGLHIA 480
QY 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAAVVISNLCVPMKE 540
DB 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAAVVISNLCVPMKE 540
QY 541 THLNERAKYVTVYQYDLELDRMSLRQHSERVLMDLGRDFCTVGLYPSCLEESPWKPPPT 600
DB 541 THLNERAKYVTVYQYDLELDRMSLRQHSERVLMDLGRDFCTVGLYPSCLEESPWKPPPT 600
QY 601 YLFSTDTGTEEFELDGEVVALPPV 623
DB 601 YLFSTDTGTEEFELDGEVVALPPV 623

```

## RESULT 4

AA042737  
ID AA042737 standard; Protein; 335 AA.

AA042737;  
08-FEB-2001 (first entry)

Human ORFX ORF2501 polypeptide sequence SEQ ID NO:5002.

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

antianeumatic gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
cholesterol ester storage; systemic lupus erythematosus; infection;  
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
bone damage; cartilage damage; antiinflammatory disease; coagulation;  
thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.  
02-APR-1999; 99US-0127636.  
05-APR-1999; 99US-0127728.  
30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;  
WPI, 2000-602362/57.  
N-PSDB; AAC76946.

Novel nucleic acids and peptides derived from open reading frame X,  
useful for treating e.g. cancers, proliferative disorders,  
neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 4178-4179; 5507P; English.

Seq	Sequence	335 AA;	50.9%;	Score 1675;	DB 21;	Length 335;
Query	Match		50.9%;	Score 1675;	DB 21;	Length 335;
	Best Local Similarity		98.8%;	Pred. No. 2.8e-147;		
	Matches 327;	Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	1	MSDQDERQINTEYAVSLLEQLKLFPEQQLFTDILIVIEGTEPPCHKVLTATCSYFRAMF				60
Db	1	MSDQDERQINTEYAVSLLEQLKLFPEQQLFTDILIVIEGTEPPCHKVLTATCSYFRAMF				60
QY	61	MSGISESQCTHYHLNVDPAITQIIITTAAYGNLAMDSTVEQLYETACFLQVEDVTLQRC				120
Db	61	MSGISESQCTHYHLNVDPAITQIIITTAAYGNLAMDSTVEQLYETACFLQVEDVTLQRC				120
QY	121	REYLIKRTNAENCYVLLSFADLLFSEELKQSKAKRVEHKFPAVYHQDAFMQLSHLLLDI				180
		REYLIKRTNAENCYVLLSFADLLFSEELKQSKAKRVEHKFPAVYHQDAFMQLSHLLLDI				180

Db	121	REYLKIKINAENCYRLISFADLFSCBEBKQSAKRVNHEKFTAVVHQDAFMQJLHDLIDI	180
Qy	181	LSNDNLNVEKEETVREAAMLMEYNTESRSQYLSSTVLSQIRIDALSEVTORAMPQGLPN	240
Db	181	LSNDNLNVEKEETVREAAMLMEYNTESRSQYLSSTVLSQIRIDALSEVTORAMPQGLPN	240
Qy	241	DKSVYVQGLYKSMPEFKPRGLGNTKEEMITFEASSENPCSLYSVCVSPQAEKYTKCS	300
Db	241	DKSVYVQGLYKSMPEFKPRGLGNTKEEMITFEASSENPCSLYSVCVSPQAEKYTKCS	300
Qy	301	PPADLHKVGTVTPDNDIYIAGGVPLKNTX	331
Db	301	PPADLHKVGTVTPDNDIYIAGGVPLKNTX	331

	RESULT	5
XX	ABB62185	
ID	ABB62185 standard; Protein; 623 AA.	
XX		
AC	ABB62185;	
XX		
DT	26-MAR-2002 (first entry)	
XX		
DZ	Drosophila melanogaster polypeptide SEQ ID NO 13347.	
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.	
XX		
OS	Drosophila melanogaster.	
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PZ	23-MAR-2001; 2001WO-US09231.	
XX		
PR	23-MAR-2000; 2000US-191637P.	
PR	11-JUL-2000; 2000US-061415O.	
XX		
PA	(PEKE ) PE CORP NY.	
PI	Venter JC, Adams M, Li FWD, Myers EW;	
XX		
WI	MPI; 2001-656860/75.	
DR	N-PSDB; ABL06288.	
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	
PT		
XX		
PS	Description; SEQ ID NO 13347; 21bp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB161675) and the encoded proteins (ABB57737-ABB72072).	
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIGO at ftp.wigo.int/pub/published_pct_sequences.	
CC		
XX		
SQ	Sequence 623 AA;	
	Query Match 17.6%; Score 578.5; DB 22; Length 623;	
Oy	Best Local Similarity 27.4%; Pred. No. 1.3e+44;	
Db	Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;	
	18 LEQLKLYEQOLFDDIVLIVEGTEFPGCHKVATLCSSYFFAMFMISGLSESGQTHTLRNV 77	
	: :	
	59 LTELNTMRHRELCDVVANVGGRIFPARHYILSCSSYFCMFGELEESRGTVTTRDI 118	

Query Match	17.6%	Score 578.5	DB 22	Length 623
Beet Local Similarity	27.4%	Pred No. 1.3e-44		
Matches 155	Conservative 103	Mismatches 215	Indels 93	Gaps 15

  

Oy	18	LEQKLFYEOOLFIDVILIVEGTEFFPCHKNVLATCCSSYFRAMNSGISSEKQTHVHLNIV	77
	:	: : : : :	:
ob	59	LTELNLMLRRHRELCDDVLANGGRIIFPHARVILSCSSYFCMFMFGELSEESQTIVTARDI	118



AAB92953  
 ID AAB92953 standard; Protein; 609 AA.  
 AC AAB92953;  
 DT 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO:11635.  
 DE Human, primer; detection; diagnosis; antisense therapy; gene therapy.  
 KM Human, primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 PN EP1074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,  
 PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length CDNA defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length CDNA -  
 PS Claim 8; SEQ ID 11635; 2537bp + CD ROM; English.  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length CDNA defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length CDNA. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length CDNA. The primers allow obtaining of the full-length  
 CC CDNA easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX  
 SQ Sequence 609 AA;  
 Query Match 17.0%; Score 558.5; DB 22; Length 609;  
 Best Local Similarity 26.1%; Pred. No. 9.1e-43;  
 Matches 150; Conservative 110; Mismatches 222; Indels 93; Gaps 14;  
 QY 9 INTEVAVSLLEOKLFYEQOLFDTIVLIVETGTEPCPKXWLVACSSYPRAMPFSGSESK 68  
 ID 46 ISDKHRCQTLIEVNLKRLKHELDVAVLVGAKKIYHRLVLSACSPFRFMFGEJLAESR 105  
 QY 69 QTHVHLRNVDAATLQIITVAYTGNLAMDSTVEQLVETACFLQVEDVLQRCREYLIKKI 128

Db 106 QTEVLRIDIDBERAMELLIDFAYTSQITVEGVNQTLIPAACTLQIAEIOACCEFLKRLQ 165  
 QY 129 NAEVCYRLISPADLPSCEELKOSAKRMVEHKTAYVHODAFMQSHDLIDILSSDNLNV 188  
 Db 166 DPNCGGIRAFADTHSCRELLRIADKFTQHNQOEWESEFMLLPANQILDISSDELNV 225  
 QY 189 EKEETREAMLMLENTESRSQVLSVLSQIRIALS-----E 227  
 Db 226 REEEQVFNAVMAMVKYSIOERRPOLPOLYQHRLPLSLKFLVGTVGSDPLKSDSECRD 285  
 QY 228 VTQRAWFOGLPPNDKSVVVOGLYKSMKPKFKRLQNTKEEMITFIASENSPCS---LYS 284  
 Db 286 LVDEAKNYLLLPQERH-LMQG-----PR-TRPRKPIRCSEVL---SAVGWCSGDALSN 334  
 QY 285 SVCYSPQAKYVKLCSPPADLHKVGTVPNDIYAGS-----QYPLKNTDNNH-- 334  
 Db 335 VERYPQTNEMRMVMSKRRCGVSVL-DLLVAVGHDGSSYLSNVERYDPTNOMS 393  
 QY 335 ---SKTSKLQTA-----FRTVCFYWFDAQONTWFEKTPMLFVRIRK 373  
 Db 394 SDMAPISTORTSVGVAVLVGFLYAVGGDGVSCINIVERYDKENKMTIRVASMSTRRLGV 453  
 QY 374 SLVCCGGYIYAIAGDSVGGELNRRRTVERDYDEKEMTVVSLPCAMQMSAAVVVHDCIYV 433  
 Db 454 AVAVLGGFLYAVGSDGTSPLN--TVERYNPQENMHTIAPMGTBRKHLGCAYQDMIVA 511  
 QY 434 M-----TLNLMYCFPPSDSVWMAMQTSRFSASAAAFDCKIFYGGLHIANNSGR 486  
 Db 512 VGRDDTTELSAERNPRTNOMSPVAMTSRRSGVLVNVGQMAVGGF----- 562  
 QY 487 LPSTGVDSGV--TVEIYDVNKNEMKMANIPAKR 519  
 Db 563 -----DGTTYLKTIEVFPDANTWRLVGMYYR 591  
 RESULT 8  
 ABB63086  
 ID ABB63086 standard; Protein; 689 AA.  
 AC ABB63086;  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polypeptide SEQ ID NO 16050.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmacological.  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmacological.  
 OS Drosophila melanogaster.  
 PN MO200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2001; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PMD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR N-PSDB; ABL07189.  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 16050; 21bp + Sequence Listing; English.  
 XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB163051), expressed DNA  
CC sequences (AB161840-AB16175) and the encoded proteins  
CC (AB161737-AB162072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 689 AA;

Query Match 16.9%; Score 555.5; DB 22; Length 689;

Best Local Similarity 27.1%; Pred. No. 2,1e-42;

Matches 168; Conservative 104; Mismatches 246; Indels 103; Gaps 22;

QY 10 NTEVAVSLIEQLKLFYEQQLFTDVLIVEGTEPCHRMVLAATCSYFRAMFSGLSKQ 69  
DB 136 NEQHTARSPDAMENMKQKQLCVILVADVEIHAHMYLASCSPYFAMFTS-FRESHQ 194  
QY 70 THVHLNVDAAATLQIIITAYTGNLAMDSTVEQLYETAKFLOVENVLQRCREYLTKIN 129  
DB 195 ARITLQSVARALELIDVYATATVENVNDNVQLLTAAVLQTLTVRACCDPLQTLQD 254  
QY 130 AENCVALLSFADLFSCCEELKQSAKRMVEKFTAVYHQAAMQSLHDLIDLSQDLNTE 189  
DB 255 AENCGLIRFADIHACVELLNTAYETIEQHFNEVIOFDEFLNLSHQVLSLGNDSISP 314  
QY 190 KEETVREAAMLWEVNTESRSQYLSVLSQIRIDALSE--VTQRAMFQGLPNDKSVVQ 247  
DB 315 NEERVVECVIAMLRYDVPMREQFTSLMEHVRLLPFLSKETIQRV-----DKETLLE 366  
QY 248 G-----LYSMKPKF-----PR--LGMKEMMFIEMSSNPSGLYSVCY 288  
DB 367 GNIVCKNLIIEALTYHLPLETKSARTVPRKPKVMK---ILLVIGQAPKAIRSEVNY 422  
QY 289 SPOAEKYYKLCSPPADLHKVGVTVTPNDIYIAGGVPLNKTNTNHSKTLQTAFTYN 348  
DB 423 DAREEKVYQAAEMPRNCRSGSLGVDK-VYAVG-----FNGS-----LRRITVD 467  
QY 349 CYWFDAGQNTWPKTPMLFVRIKPSLVCCGYIYVAGISVGGELNRRTEREYDEKDE 408  
DB 468 V---YDPATDQVANCMEARRETLGVAVLNGCIYAVG--FDGTGLSAAEWYDKTDI 522  
QY 409 WTMVSLPCAMQWQAIVVYHDCIYV-----TLNLMKCYFPSPSDSWEMAMQTR 459  
DB 523 WRFIASMSTRSSVGVVHGLLYAVGVDGFTROCLSSVERNPDTDMVNVAAEMSSRR 582  
QY 460 SPASAAAFGDKIFYIGLHIATNSGIRLPSGTVDGSSV--TVETIYVNNKEMMANIPA 517  
DB 583 SGAGVGVANNILYAVGG-H-----DGMVRRSEAYICETNSMSVADMSY 627  
QY 518 KQYSDPCVAVVINS--LCVFMRETHLNERAKYVTYQVLELDMSLRQHSERVLMDL 575  
DB 628 CRRN-----AGVVAHGDGLYVVGDDGTSNLASVEYVCPD--SDSW-----RLPL 672  
QY 576 GRDFCTVCKLTPS-CLEESP 595  
DB 673 -----WTIGRSYAVGCMIDKP 688

RESULT 9  
ABB90399  
ID ABB90399 standard; Proteoin; 606 AA.  
AC ABB90399;  
XX  
XX  
XX 24-MAY-2002 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 2775.  
XX

KW Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
OS Homo sapiens.  
PN WO200190304-A2.  
XX  
XX  
XX 29-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-US16450.  
XX  
XX 19-MAY-2000; 2000US-205515P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Birse CE, Rosen CA;  
XX WPI; 2002-122018/16.  
XX DR N-PSDB; ABL90808.  
XX

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
PS Claim 11; SEQ ID NO 2775; 2081bp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 606 AA;

Query Match 16.8%; Score 554; DB 23; Length 606;

Best Local Similarity 26.6%; Pred. No. 2,4e-42;

Matches 155; Conservative 93; Mismatches 240; Indels 94; Gaps 13;

QY 6 EROINTE---YAVSLIEQ-LKLFYEQQLFTDVLIVEGTEPCHRMVLAATCSYFRAMF 61  
DB 4 QRELAELRLVSTLQOGGLKDLDBKFTICTLTAAGSKSPCHNLIISACSPYRREYVL 63  
QY 62 SGLSSKQTHVHLNVDAAATLQIIITAYTGNLAMDSTVEQLYETAKFLOVEDYVLRQR 121  
DB 64 SVIDAKKEVVLVDVDPALIDLILIKYLSASIDLDNGVODIFALASRFQIPSYFTVCV 123  
QY 122 EYLKIKIANGCVRLLSFADLFSCCEELKQSAKRMVEKFTAVYHQAAMQSLHDLIDL 181  
DB 124 STLQRLAPGNCLALIRGLLDCPRLAISAEFVSDRFVQCKEDFWQLSPQELISVI 183  
QY 182 SSDNLNVEKETVREAAMLWEVNTESRSQYLSVLSQIRIDALSE----- 227  
DB 184 SNDSLNVEKEAVFPAVWKVWRTDCKENVKQLSEVFDICIRPLMEKPKHVEKODIIX 243  
QY 228 -----VTQRAMFQGLPNDKSVVVGGLYK-----SMKPKF--PRIGMTE 266

Db 244 SNPDLOKKIKVLDKAPAGKLPBESKNAKTGAGEVNGDVEDDLPGYLANDIPRHGNEVK 303  
Qy 267 EMMIFIBASSENPCSLYSVCYSPQAEKVVYKLCSPADLHKVGTVTVPNDIYIAGQVP 326  
Db 304 DLILVNDT-----AAVAYDTENECVLTALAEIPRNHSSIVYQOQIYVGGLY- 354  
Qy 327 LKQTKTNHSTKSKLQTAFTFVNCVYFWDAAQNTWFPPTPLFVRIKPSLVCCRGY----- 361  
Db 355 -----VDEENKDPQLQ-----SYFFQLDLSASEWVGPLP-----PSARCLFGLGEVDD 398  
Qy 382 -IYALIGSDSVGELNRTVRYDEKCEKEMTVSPCLPCAMQMSAAVVVHDDIYV----- 434  
Db 399 KIYVNAKDJQTEASLDVLCYDPVAAKMEVKKLPKTVYGHNVISHKMTYCLGCKTDD 458  
Qy 435 --TLNLMYCYFPPSDSVWEMAMRQTSRSPASAAFGDKIFYIGLHATNSGIRLPSGTV 492  
Db 459 KKCINRVFIINPKKGDMDLAPMKIPRSMFGVAVHKKIYIAGV-----TE 505  
Qy 493 DGSSTVEIYDVNKNEMKMANIPAKRYSDPCYRAVVISNSL 534  
Db 506 DGLSASVEAFDLITNKMDVMTFPOERS--ISLVLAGSL 544

## RESULT 10

AAM38956 standard; Protein; 587 AA.

AAM38956;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2101.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;  
leukemia; thrombolytic; drug screening; arthritis; inflammation;

Homo sapiens.

MO20015312-A1.

26-JUL-2001.

26-DEC-2000; 2000MO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

WPI; 2001-442253/47.

N-PSDB; AAI58112.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

The invention relates to human nucleic acids (AA157798-AA161369) and

the encoded polypeptides (AAM38642-AA42213) with noctropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localized neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilization of the activities such as: Immune system suppression,  
CC actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

SQ Sequence 587 AA;

Query Match 16.4%; Score 541.5; DB 22; Length 587;

Best Local Similarity 25.9%; Pred. No. 3.3e-41;

Matches 156; Conservative 106; Mismatches 216; Indels 125; Gaps 19;

Qy 14 AVSLLEQLKLFYEQOLFDTIVLIVGTEPPCHKQVLAATCSYFRAMFGSLSESKOTYH 73  
Db 36 AFKVMNEER--SKQLLQDWTVAEDVEIAHRVLAACSPYFCAMFTGDMSESKAKTE 92  
Qy 74 LRVNDAATLQIIITVAVTGNLAMDSTVEQLYETACFLOVEDVLRQCREYLKIKINAENC 133  
Db 93 IKVDQGLSKLIDYITVAIEVTEENQVLLPASLILQMDVRQCCOFLQSLHPITNC 152  
Qy 134 VRLISFADLPSCBELKQSAKRVNHFVAVYHQAQFMQLSHLLIDILSSDNLNVEKET 193  
Db 153 LGIRAFADVHTCTDLQCANAAVEQHFEVWLGEEFLSLDQVCSLSSDXLTYSSEK 212  
Qy 194 VBEAAMLMLEYTESRSQVLSQIRDALSEVTQRAWFQGLPNDSSVVVQGLYKSM 253  
Db 213 VBEAIVSWINYEKENLEMAKLEHVLPL-----PRD--YVQ----- 251  
Qy 254 PRFPRKRLQMTKEEMMIFIBASSENPCSLYSVCYSPQAEKVVYKLCSPADLH----- 306  
Db 252 -----TVEEBALI-----KNNNTCK-----DFLEAMKYHL---PLDQRLLIKNDP 290  
Qy 307 -KVGIVYTPNDIYIAGQVPLKNTKTNHSTKSKLQTAFTFVNCVYFWDAAQNTWFPPTP 365  
Db 291 TKRTPFVSLPKVMIVVGGQAP-----KIRSVCC--YDFEEDRMDQIAE 332  
Qy 366 MLFVRIKPSLVCCGTYAIGSDSVGELNRTVRYDEKCEKEMTVSPCLPCAMQMSAAV 425  
Db 333 LPSRRCRAGVFMAGVITAVGG--FNGSLRVRTVYDGVKQMTSISMQERRSTLGA 390  
Qy 426 VVHDCIYVW-----TLNLMYCYFPPSDSVWEMAMRQTSRSPASAAFGDKIFYIG-- 476  
Db 391 VANDLLIYAVGFDGSLASVEAYGKTNEMFVAPMNTROS VGVVGGGLYAVGVD 450  
Qy 477 -----LHIA-----TNSGIRLPSGTV-----DGSV--TWEL 502  
Db 451 GASROCLSTVEQYNPATNEMIVYADMTSRSGAGVSGQLYATGHHGDLVRKSVEVY 510  
Qy 503 DVNKNEMKMANIPAKRYSDPCVR--AAVVISLQVPMRETHLNRKAVYVYQYLEDR 560  
Db 511 DGTNTMKVADM-----NMCRRNAGCAVGLIYVGGDDGSCNLASVEY--YNPVTDK 563  
Qy 561 WSL 563  
Db 564 WTL 566

## RESULT 11

AAM78511 standard; Protein; 593 AA.

AAM78511;

06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1173.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
OS Homo sapiens.  
XX  
XX MO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
PF  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAKS1644.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 3416-3417; 6221pp; English.  
PS  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 593 AA;  
SQ  
Query Match 16.2%; Score 534; DB 22; Length 593;  
Best Local Similarity 25.3%; Pred. No. 1,7e+40;  
Matches 138; Conservative 100; Mismatches 199; Indels 108; Gaps 14;  
QY 5 DERQINTE-----YAVSLLEQLKLFYEQQLFTDIVLIVEGTEFPCHKVLAAT 51  
DB 20 DSKDNDTEHGCPVTYVPMHMKKAFKVMNELR--SQNLICDVTIYVEMDEISHRVVLA 76  
QY 52 GSSTPRAMMSGLSEKQTHVHLRNDATLTQIITTYANTGNLAMDSTVEQLYETACPL 111  
DB 77 CSPTPRAMFTGMSERARVRKEVDGWTLRMLIDYVTAIQVTEENVQVLLPAAAGLL 136  
QY 112 QVEDVLPQRCREYLKIKINACVRLSPADLFSCEELTKQSAKRMVHKFTAYVHODAFMQ 171  
DB 137 QDQDVYKTCERPLESGQLHVNCLGIRAFDMHACTDLMLKANTYAEQHFADVVLSEEFIN 196  
QY 172 LSHDLLIDILSSDNLNVEKEETVREAAMLMLEYNTESRQQLSSVLSQIRIDALSE--VT 229

DB 197 LGIEQVCSLISSDKLTISSEKVFPAVIAVWHDQVQOEPMARIMHVRPLPLPREYLV 256  
QY 230 ORAMWQGLPND---KSVVVOGL--YKSMKPFKRLMTGKEMMIPIBASNNPCLSVSS 285  
DB 257 QVEEBALVKNSSACKNVLIEAMKYLHP-----TEOKITL----- 292  
QY 286 VCSYSPQAEKVYLLCSPPADLHKVGTVPVNDIYAGQVPLKNTKTNKSKTKLQTAFR 345  
DB 293 -----KSVTRLRLTPMNLPLKLMVV-----GQAP-----KAIR 321  
QY 346 TYNCFYWPDAQNTWPKTPMLFVRIKPSLYCCBGYIYAIGDSVGBLNRRTVERDYTE 405  
DB 322 SVEC---YDFKEQRHQAVALPSRRCRAGMYMAGLVPAVGQ--FNQSLRVTVDSYDPV 376  
QY 406 KDEMTNVSPLPCAMQMSAAVVHDCIYV-----TINLWCFPRSDSWEMAMRQTS 458  
DB 377 KDQMTSVAMNRDRRSTLGAALVINGLLVAVGFDGSTGLSSVEAYNIKSNEMFHVAPMNR 436  
QY 459 RSPASAAAFGDKIFYIGLHATNSGIRLPSTVDGSS-----VTVEIYDVNENKMMAN 514  
DB 437 RSSVGVGWGLLVAVG-----YDGA8ROYLSTVECYNATTNENTYIAE 481  
QY 515 IPAKR 519  
DB 482 MSTRR 486  
RESULT 12  
AAM79495  
ID AAM79495 standard; Protein; 599 AA.  
XX  
XX AAM79495;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human protein SEQ ID NO 3141.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
OS Homo sapiens.  
XX  
XX MO200157190-A2.  
PN  
XX  
XX 09-AUG-2001;  
PD  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
PF  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52628.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 264; 6221pp; English.

XX The invention relates to polypeptides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM60302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM6020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 XX Sequence 599 AA;  
 Query Match 16.1%; Score 530; DB 22; Length 599;  
 Best Local Similarity 25.1%; Pred. No. 4e-40;  
 Matches 137; Conservative 100; Mismatches 200; Indels 108; Gaps 14;  
 QY 5 DEQINTE-----YAVSLLEQLKLFYEQQLFTDVLIVEGTEFPCHKVLAAT 51  
 DB 26 DSKDDTEKHCPTVTPMNMKKAFKYNNELR---SQNLCDVTIVADMEISARVYLA 82  
 QY CSSTFRAMFMSGLSESQKTHVHLRVNDAATLQIITTAATGNLAMDSTVEQLYETACFL 111  
 DB 83 CSSTFRAMFMSGLSESQKTHVHLRVNDAATLQIITTAATGNLAMDSTVEQLYETACFL 142  
 QY 112 QVEDVLQRCREYLIIKKINENCYRLLSFADLFSCBELKQSAKRVBEHKFTAVYHODAFMO 171  
 DB 143 QLDGVKTCCEFLSEQLHPNCGIRAFADMACTDLNKAANYAEQHFADVLLSEEPFN 202  
 QY 172 LSHDLIDILSSNNLVEKETVREAAMLLENTSRSQYLSVLSQIRIDALSE--VT 229  
 DB 203 LGIEQVCSLISDKLTISSEKVEAVIAVNNHDKVROFMAKLMHEVRLPLPREYLV 262  
 QY 230 QRAMFQGLPND---KSVVQGL-YKSMPEFKRLGTEEMWIFLEASSENPCSLYS 285  
 DB 263 QRVEBALVNSACKNYLLEAKKHLHP-----TEQRILM----- 298  
 QY 286 VCYSPOAEKYYKLCSPADLHKVGTVPNDIYIAGGVPLNKTNTNSKTSKLOTAPR 345  
 DB 299 -----KSVRTLRTPMNLKLMVVY-----GQGNP-----KALR 327  
 QY 346 TVNCFWFDQAQNTWPKTMLFVRIKPSLYCEGTYIAGDSVGGELNRRPTVERKYDT 405  
 DB 328 SAEC---YDFKEGRHQVAFELPSRRCRAGNYLAGLVFAVG--FNGSLRVRITVDSYDPV 382  
 QY 406 KDEMTWVSLPCAMQMSAAVVVHDCIYVM-----TLNLMCYCFPRSDWVEMAROTS 458  
 DB 383 KQMTSVANMRDRSTLGAVALNGLLYAVGFDGSGLSVEAYNKKSNEMFVAAPMNR 442  
 QY 459 RSFASAAAFGDKI FYIGGLHIAITNSGIRLPSGTVDSS-----VTVEIYDVKNEWMKMAAN 514  
 DB 443 RSSVGVGVGGLLYAVG-----YDGNASROYLSTVCYNAATNTMETYIAYE 487  
 QY 515 IPAKR 519  
 DB 488 MSTRR 492

RESULT 13  
 ID ABB60307 standard; Protein; 744 AA.  
 AC ABB60307;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Drosophila melanogaster polypeptide SEQ ID NO 7713.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 XX  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI; 2001-656660/75.  
 DR N-PSDB; ABL04410.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Disclosure; SEQ ID NO 7713; 21pp + Sequence Listing; English.  
 PS  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01940-ABL6175) and the encoded DNA  
 CC sequences (ABBS7737-ABBS7072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 744 AA;  
 Query Match 16.0%; Score 526.5; DB 22; Length 744;  
 Best Local Similarity 28.0%; Pred. No. 1.2e-39;  
 Matches 157; Conservative 77; Mismatches 207; Indels 119; Gaps 18;  
 QY 11 TEVAVSLLEQLKLFYEQQLFTDVLIVEGTEFPCHKVLAATCSYFRAMFMSGLSESQK 70  
 DB 38 SNVAKBALKMMYMRSHGKLTIDVLEVKKELPRAKVVLSAASPYFKAMFTGGLSESS 97  
 QY 71 HVHLRVNDAATLQIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRCREYLIIKKINA 130  
 DB 98 RQLOGVCTANRSRLIFYNTQIIVTEVTQQLPAATMFPVNVINDCCAFLEGRQDP 157  
 QY 131 ENCVRLLSFADLFSCBELKQSAKRVBEHKFTAVYHODAFMOQLSHDLLIDILSSNNLVEK 190  
 DB 158 TNAIGIAHFAEQHGVEQLQKANAIFERNFTQVCOEEBELQLSAVQALIALIRDELINVOE 217  
 QY 191 EETVEAAMLLENTSRSQYLSVLSQIRIDALSEVTQRAMFQGLPNDSSVVVQGLY 250  
 DB 218 EREVYNAVLKWKVYEDNHNKCMHEHILGAVRC-----OPLTPN----- 255  
 QY 251 KSMPEFKRLGTEEMWIFLEASSENPCSLYSV--CYSPOAEKYYKLCSPADLHKV 308  
 DB 256 -----FLKEQM-----KNCDVLRVPACREYLA-KIRF-----DLTLHKC 289  
 QY 309 GTVV--TPDND--IYIAGG-----QVPLNKTNTNSKTSKLOTAP-- 344  
 DB 290 PGVKERTPTWTMIFVAGGFRRHSIDLIAVYVDDMTWTTLANLRI---PRSGLGAAPFK 346  
 QY 345 -----RTVNCFY--WFD--AQONTMFPKTPMLFVRIKPSLYCEGTYIAGDS 388  
 DB 347 GRFYAVGGRNNNIGSSYDSWDVDRYSVATETWRPCAPMSVPHRRVGVAVDELMYAVG- 405



Qy 389 SYGGEINRRTVERDYTEKDEWTVSPLPCAMQMSAAVVVEDCIYVW-----TLNLMYC 441  
 Db 406 SAGMEYH-NTVEYDDELDLRMTLVQPMHAKRLGVGVVVMRLLYAIGFCDGERLASVVC 464  
 Qy 442 YPRPSDWMEMAMRQSRSPASAAFGDKIFYIGLHININSIGIRLPSGTVDSS--YTV 499  
 Db 465 YHPENNEMSFLLPPLQGRSGAGVAAINQYIVVGGF-----DETROLATV 509  
 Qy 500 EYDVNKNEMKMAANIPAKR 519  
 Db 510 ERYDTENDTMDVAPIQIAR 529

## RESULT 14

AAM47944  
 ID AAM47944 standard; Protein; 575 AA.

AC AAM47944;

DT 01-MAR-2002 (first entry)

DE Human viscentric cyclobutulin protein 63.

KM Human; viscentric cyclobutulin protein 63; malignant tumour; haemopathy;  
 KW development confusion disease; human immunodeficiency virus; HIV;  
 KW infection; immune disease; inflammation.

OS Homo sapiens.

PN CN1311211-A.

PD 05-SEP-2001.

PF 02-MAR-2000; 2000CN-011811.

PR 02-MAR-2000; 2000CN-011811.

PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

PI Mao Y, Xie Y;

DR WPI; 2002-049907/07.

DR N-PSDB; ABA05463.

PT New human cyclobutulin 63 polypeptide and encoding polynucleotide useful  
 for treating tumour, haemopathy and human immunodeficiency virus -  
 PS Claim 1, Page 27-28 (Disclosure); 35pp; Chinese.

CC The invention relates to human viscentric cyclobutulin protein 63, its  
 CC recombinant production, antagonist, encoding polynucleotide and  
 CC application. The polypeptide is useful for treating malignant tumour,  
 CC haemopathy, development confusion disease, human immunodeficiency virus  
 CC infection, immune disease and various inflammations.

CC Sequence 575 AA;

Query Match 15.9%; Score 523; DB 23; Length 575;

Best Local Similarity 25.0%; Pred. No. 1.7e-39; Indels 114; Gaps 15;

Matches 151; Conservative 101; Mismatches 237; Indels 114; Gaps 15;

Qy 2 STQDERQINTEYAVSLLEQLFYEQQLFTDIYLVGTEPCKHNVLATGSSYFRAMF 61

Db 10 SSSSEFYQVNHAEQFRKMGESYLKQOQCDVILIGNKRIIPAHRLVLSVSDYFAMFT 69

Qy 62 SGLSEKQTHVHLRNVDAAITLQIITYAYTGNLANDSYVEQLYETACFLQVEDVLRGR 121

Db 70 SYVCEAKQKEIKMEGIDPALMDLVQFAATGCELEKEDTLENLAACLLQ.PQVVEVC 129

Qy 122 EYLTKKINENCRILSPADLPSCCELKQSAKMVHNKTAAYVHQDAFQLSHDLIDL 181

Db 130 HFLMKLHPNSCLGIRAPADAGCIELMKVAISYTMENINEYVRNDEPILLPAEELHKL 189

Qy 182 SSDNINVEKETREAAAMLMEYTESSSQYLSVLSQIRI-----DALSEYTORAMFG 236  
 Db 190 ASDDVNVEDEETIFPALMMVXYDMQSRCDLSMLAFIRLPLPPQLADLENHALFKN 249  
 Qy 237 LPBNKSVVQGLYKSMK-----FFKR----- 260  
 Db 250 DLECKLILAMKHYLLPERRTLMQSPRTKPKSTVGTLYAVGMDNNKGAATTIERYDLR 309  
 Qy 261 -----LGMTKEEMMIFTEASENPSYSSVCSYSPQAKYKLCSPAD 304  
 Db 310 TNLMTQAGMMNRRLOPGVAVIDDLKLVYIGRDGKLTANTVECTNPKTKTWVL--PPMS 367  
 Qy 305 LHKVGTVP--DNDIYIAGQVPLNKTNTNKSITSKLOTAFTVNCFYWPAQONTPEPK 363  
 Db 368 THRHGIGTVLEGPYAVVG-----HDGMSYLVNTER-----WDPOQOQWTFV 410  
 Qy 364 TPLMFLVRILKPSLVCCGITYAIGDSVGGELNRRIVERDYDEKDEWTVSPLPCAMQMSA 423  
 Db 411 ASMSIARSTVGVALNGKLYSVGRD--GSSCLSMSEYDPTNTNMNCAPM-CGRGCV 467  
 Qy 424 AVVVD-CIYVW-----TLNLMCYPRPSDWMEMAMRQSRSPASAAFGD 469  
 Db 468 GVATCDGFLYAVGVDAPASNHCSRLDYVERDYDEKDTWTVAFLSMPRAVGVCLGSD 527  
 Qy 470 KIFYIGLHIAATNSGIRLPSGTVDSSV--TVEYDVNKNEMKMAANIPAKRYSDPCVRA 527  
 Db 528 RLAVGG-----YDQGYLTMTMESYDPTNEMTQMAALNIGR-AGACV-- 569  
 Qy 528 VVI 530  
 Db 570 VVI 572.

## RESULT 15

AAM40225  
 ID AAM40225 standard; Protein; 748 AA.

AC AAM40225;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3370.

KM Human; nocrotic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Auand V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AU, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dmanac RT;

DR WPI; 2001-442253/47.  
DR N-PSDB; AA159381.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX  
PS Example 5; SEQ ID NO 3370; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA038642-AA042213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX

Sequence 748 AA;

Query Match 15.9%; Score 523; DB 22; Length 748;

Best Local Similarity 25.0%; Pred. No. 2.6e-39;

Matches 151; Conservative 101; Mismatches 237; Indels 114; Gaps 15;

QY 2 STDERQINTYAVSLLEQDLFEYQQLFTDILVIGTEFPCHKVLAATCSSYFRAMFM 61  
DB 183 SSSEEFYQAVHABQFRKMSYTLKQQLCDVILVGNRKIPARLVLSSVSDYFAMFT 242  
QY 62 SGSESQOTVHLRNDVATLQIITAYTGNLANMDSTYEOUYETRAFLQVEDVULQRCR 121  
DB 243 SDVCEAKQOEIKMEGIDPMLMDLVQFAYTGCLELKEDITENLAAACLLQLPQVVEVCC 302  
QY 122 EYLTKKINAENCVRLLSFADLFCEELKQSAKRWHEKFTAVYHODAFMQLSHDLLIDIL 181  
DB 303 HFLMKLIHPNSCGIRAFADAGCIELMKVASHYTMENIMEVIRNOEFLLPRAEELHKL 362  
QY 182 SSDNLVKEKETVREAAMLMLEYNTESRQYLSVLQIRI-----DALSEVTQRAFPQ 236  
DB 363 ASDDVVVPDEETIFHALMMWVKYDMQSRCDLSMLAFIRLPLPQILADLENHALFKV 422  
QY 237 LPPNDKSVVQGLYKSPK-----FFPR----- 260  
DB 423 DLSCQKILIEAMKYHLPERRTLMQSPRTKPKSTVGTLYAVGMDNNKGATTIEKYDLR 482  
QY 261 -----LGMTKEEMMTFIEASSENPCSLVSSVCYSPQAEKVYKLCSPAD 304  
DB 483 TNLMIQAGMNGRRLQPGVAVIDDKLFVIGRGDLKTLNTVECTYNEKTKTWVL--PMS 540  
QY 305 LHKVGTVPD-DNDIYIAGQVPLKNTKTNHSTKSLQTAFTVNCYWFDAQOQNTWPK 363  
DB 541 THRHGLGVTLLEGPIYAVG-----HDCMSYLTNVER-----WDPSQQWTFV 583  
QY 364 TPMLFRIKSLVCEGYIYAIGDSVGEHLNRRIVERTYEDKEDWTMVSPLPCANQMSA 423  
DB 584 ASMSIARSTVGAALNGKLYSVGRD--GSSCLSSMEYDPTNKNMCAFM-CKRRGV 640  
QY 424 AVVVDH-CIYV-----TLNLYCYFPRSDSVEMAMQTSRSPASAAAFD 469  
DB 641 GVAICGFLYAVGHDAPASNHCSRLLDYERDKTDITWTWVAPLSMPRDANGVCLLD 700  
QY 470 KIFYIGLHATNSGIRLPSGTVDGSSV--TVEIYDNKNEMKMAANIPAKRYSDPCVRA 527  
DB 701 RLAVAVG-----YDGGTYLNTMESYDPQTNEMTQMASLNIQR--AGACV-- 742  
QY 528 VVI 530  
|||

DB 743 VVI 745  
Search completed: July 14, 2003, 18:14:54  
Job time : 24.5998 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 ; Search time 13.7827 Seconds  
(without alignments)  
4345.419 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292  
Sequence: 1 MSTODERQINTXAVSLLEQ.....STDGTEEFELDGEWALPPV 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	57.3	348	2 T17304	hypothetical prote
2	559	17.0	1476	2 A45773	hypothetical prote
3	497.5	15.1	531	2 T33099	hypothetical prote
4	438	13.3	618	2 T33222	kelch-like protein
5	415	12.6	836	2 T33222	hypothetical prote
6	373.5	11.3	625	2 T25373	hypothetical prote
7	327	9.9	564	2 C42523	A55R protein - vac
8	324	9.8	564	2 J01792	Salp17R protein -
9	312.5	9.5	559	2 S24675	p65 protein - Ect
10	292	8.9	585	2 T37216	calicin - human (f
11	284	8.6	588	2 T46003	calicin - bovine
12	244.5	7.4	509	1 MVMZMX	M9-R protein - myx
13	235.5	7.2	817	2 T24063	hypothetical prote
14	232	7.0	512	1 MZVZB6	59K HindIII-C prot
15	227	6.9	512	2 C42504	C2L protein - vacc
16	226	6.9	579	2 H88478	protein F47D12.7 l
17	216	6.6	480	2 G36213	F3L protein - vacc
18	209	6.3	476	2 T30781	kelch protein - vacc
19	208	6.3	480	2 H42506	F3L protein - vacc
20	203	6.2	251	2 C36418	M-T8 protein - myx
21	185	5.6	552	2 T54388	L2TR-1 - human
22	183.5	5.6	518	2 S42387	MTP protein homol
23	182.5	5.5	581	2 T31745	hypothetical prote
24	178.5	5.4	274	2 S44616	C08C3.2 protein -
25	177.5	5.4	524	2 JC7090	germ cell-less pro
26	175.5	5.3	424	2 T39310	zinc finger protei
27	175.5	5.3	602	2 T21980	hypothetical prote
28	171.5	5.2	586	2 D84710	hypothetical prote
29	170.5	5.2	794	2 S59069	Z13 protein - mous

30	169.5	5.1	480	2 T20168	hypothetical prote
31	168	5.1	570	2 T24152	hypothetical prote
32	168	5.1	1466	2 T32791	hypothetical prote
33	166.5	5.1	410	2 T15442	hypothetical prote
34	166	5.0	320	2 T27372	hypothetical prote
35	166	5.0	484	2 T39200	zinc finger protei
36	162.5	4.9	410	2 S44626	C50C3.8 protein -
37	158	4.8	320	2 T31948	hypothetical prote
38	154	4.7	326	2 E86351	hypothetical prote
39	153.5	4.7	736	2 T00023	transcription fact
40	152.5	4.6	202	2 S16442	MTP protein - mou
41	152	4.6	551	2 T50663	RCCL-like G exchan
42	151	4.6	242	2 T31950	hypothetical prote
43	151	4.6	301	2 T33335	hypothetical prote
44	149.5	4.5	301	2 T32803	hypothetical prote
45	148	4.5	290	2 T32096	hypothetical prote

## ALIGNMENTS

RESULT 1  
T17304  
hypothetical protein DKFP566C134.1 - human (fragment)  
C/Spectrum: Homo sapiens (man)  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 15-Oct-1999  
C/Accession: T17304  
R: Bloembergen, H.; Boeckler, M.; Brandt, P.; Mewes, H.W.; Gasteiger, J.; Wilmann, S.  
submitted to the Protein Sequence Database, September 1999  
A/Reference number: Z18724  
A/Accession: T17304  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-348 <BIO>  
A/Cross-references: EMBL:AL117562  
A/Experimental source: fetal kidney; clone DKFP566C134  
C/Genetics:  
A/Note: DKFP566C134.1

Query Match	57.3%	Score 1886	DB 2	Length 348
Best Local Similarity	99.4%	Pred. No. 2.4e-140		
Matches 346	Conservative 1	Mismatches 1	Indels 0	Gaps 0
QY	276	SENPCSLYSVCYSPQAEKRYKLCSPADLHKVGVTPDNDIYIAGGVPLNTKTNHS	335	
DB	1	AENPCSLYSVCYSPQAEKRYKLCSPADLHKVGVTPDNDIYIAGGVPLNTKTNHS	60	
QY	336	KTSKLQTAFTVNCYFWDQNTWPKTPMLFVRIKPSLVCCGGYIYVAGDSVGGELN	395	
DB	61	KTSKLQTAFTVNCYFWDQNTWPKTPMLFVRIKPSLVCCGGYIYVAGDSVGGELN	120	
QY	396	RTVERVYDEKEMTVSPLPCAKOMSAVVVHDCYVNTLMLKCYPPRSQWVMAR	455	
DB	121	RTVERVYDEKEMTVSPLPCAKOMSAVVVHDCYVNTLMLKCYPPRSQWVMAR	180	
QY	456	QTSRSPASAAAGDKIFYGGLHIAITNSGIRLPSGTVSSVTVBIYDVKMEKKAANI	515	
DB	181	QTSRSPASAAAGDKIFYGGLHIAITNSGIRLPSGTVSSVTVBIYDVKMEKKAANI	240	
QY	516	PAKRYSDPCVRAVVISNLSICVFMRETHLNERAKYVTVQYDLELDRLWSLRQHSERVLMDL	575	
DB	241	PAKRYSDPCVRAVVISNLSICVFMRETHSNERAKYVTVQYDLELDRLWSLRQHSERVLMDL	300	
QY	576	GSDPRCTGVKLYPSCLESBPMKPPYLPSTDGTEBEELDGEWALPPV	623	
DB	301	GSDPRCTGVKLYPSCLESBPMKPPYLPSTDGTEBEELDGEWALPPV	348	

RESULT 2  
A45773  
kelch protein, long form - fruit fly (Drosophila melanogaster)  
N/Alternate names: ring canal protein kelch  
N/Contains: kelch protein, short form

C/Species: Drosophila melanogaster  
C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 22-Oct-2001  
C/Accession: A45773  
R/Xue, F.; Cooley, L.  
Cell 72, 681-693, 1993  
A/Title: Kelch encodes a component of intercellular bridges in Drosophila egg chambers.  
A/Accession: A45773, MUID:93201592; PMID:8453663  
A/Reference number: A45773  
A/Molecule type: mRNA  
A/Residues: 1-1476 <XUE>  
A/Cross-references: GB:108483  
A/Experimental source: embryo  
A/Note: Sequence extracted from NCBI backbone (NCBIN:127754, NCBIPI:127756)  
C/Comment: This long form is produced by partial suppression of a TGA codon by an unknown C/Genetic:  
A/Gene: FlyBase:kel  
A/Cross-references: FlyBase:FBgn001301  
C/Superfamily: POZ domain homology  
F/1-1476/Product: kelch protein, long form #status predicted <KPL>  
F/1-688/Product: kelch protein, short form #status predicted <KPS>  
F/142-244/Domain: POZ domain homology <POZ>  
F/689/Region: opal stop codon

Query Match 17.0%; Score 559; DB 2; Length 1476;  
Best Local Similarity 27.1%; Pred. No. 4e-35;  
Matches 168; Conservative 104; Mismatches 250; Indels 98; Gaps 22;

10 NTEYAVSLLEQQLFYEQQLFTDIVLIVGTEFPCHQVLTATCSYFRAMFMSGLSESKQ 69  
135 NEQHTARSPAMMBEMKQKQOLCDVILADVDETHAHMVLASCSFYFAFTS-FEESRQ 193  
70 THVHLNVDATLQIIITVAYTGNLAMDSTVEQLYETACFLQVEDVLQRCREYLIIKIN 129  
194 ARTLQGVDAARALBELLDVYTTATVEVNEEDVQLLTAANLLQLTVDRADCCFLQQLD 253  
130 AENCVRLLSPADLFSCBELKQSAKRMVHEKFTAVHODAFMQLSHDLIDILSSDNLNVE 189  
254 ASACLGRFADPHACVLELNVATYEQHFNEVIGPDEFNLNHEQVLSLINDRISIV 313  
190 KEETVEAAMLMLEYNTESSQVLSVLSQIRIDALSE--VTORAMQGLPRNDKVVVQ 247  
314 NERRVVECVAMLRVDPMBEOSTLMEHVRLEPFSKEYITQV-----DKETLLE 365  
248 G-----LYKSMPEKFK-----PR--LGMTKEEMMIFIEASSENPCSLYSVCY 288  
366 GNIVCKNLIIEALTYHLPTETKSARTVPRKPGVMPK-----ILLVGGQAPKAIKRSWEV 421  
289 SPAEKYKLCSPPADLHKVTVTPNDIYIAGGOVPLKNTKTNHSTKSLQTAFTVN 348  
422 DLREKKYQAEMPMNRCSGLSVLGDK-VYAVGG-----FNGS-----LRVTV 466  
349 CFYMPDAQONTWPKTLMFVRIRKPSLVCEGYIYAIGDSVGGELNRRTERVYDEK 408  
467 V-----YDPAIDQMANCSMEARRSTLGVAAINGCIYAGG--PDGTTGLSAAEMVDPDT 521  
409 WTVVSPPLPCAMQWSAAVVVHDCIYV-----TLNLMYCYFPRSDSVEMAMQTSR 459  
522 WRIASSTRSSVGVVGLIYAVGGVDFTRQCLSSVERYNPDITVNVAAEMSSR 581  
460 SFPSAALFGDKITFYIGGLIATNSGIRLPSGTVDGSSV--TVEIYVNVKNEWMAANIPA 517  
582 SGAGVGLNNILYRVG-H-----DGPVRSVEAYCETMSRMSVADMSY 626  
518 KRVSDPCVRAVVISNS--LCVPMRETHLNRAKYVYQVLEDRWSLRHISERVLMDL 575  
627 CRNN-----AGVAHGLLTVVGDDGTSVLASVEVYCPD--SDSWRLP-----ALMTI 674  
QY 576 GRDPR--CTVQKLYPSCLER 593  
DB 675 GREYAGVCMIDK--PKMEE 692

RESULT 3

T33099  
hypothetical protein R12E2.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 28-Jul-2000  
C/Accession: T33099  
R/Goela, D.; Scheet, P.  
Submitted to the EMBL Data Library, May 1998  
A/Description: The sequence of C. elegans cosmid R12E2.  
A/Reference number: Z21281  
A/Accession: T33099  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-531 <GOE>  
A/Cross-references: EMBL:AF067219; PIDN:AA17022.1; GSPDB:GN00019; CESP:R12E2.1  
A/Experimental source: strain Bristol N2; clone R12E2  
C/Genetics:  
A/Gene: CESP:R12E2.1  
A/Map position: 1  
A/Intons: 34/3; 60/3; 95/1; 186/3; 240/3  
C/Superfamily: POZ domain homology  
F/42-145/Domain: POZ domain homology <POZ>

Query Match 15.1%; Score 497.5; DB 2; Length 531;  
Best Local Similarity 26.5%; Pred. No. 5.7e-31;  
Matches 134; Conservative 82; Mismatches 226; Indels 63; Gaps 10;

10 NTEYAVSLLEQQLFYEQQLFTDIVLIVGTEFPCHQVLTATCSYFRAMFMSGLSESKQ 69  
35 NSHFANSVLIQQLGKLRVLCVILICGKRNHNRVLSLCSFPLSMFSQMAECIM 94  
70 THVHLNVDATLQIIITVAYTGNLAMDSTVEQLYETACFLQVEDVLQRCREYLIIKIN 129  
95 REINNEIRPITLALIEFCYTGALIDDSNVODILPAACLIQIHEVQACCDYLKKQD 154  
130 AENCVRLLSPADLFSCBELKQSAKRMVHEKFTAVHODAFMQLSHDLIDILSSDNLNVE 189  
155 PNCIGIRAFADTHCKELSSADEFALKNFSRVIGKEFQMLVLESITITIKSDKLNA 214  
190 KEETVEAAMLMLEYNTESSQVLSVLSQIRID-----ALSEVTORAMFQGLPRNDKS 243  
215 SEELVPSAVIQVNRHDSIKRTHLSMLSHVLPCTKPELVSVSEELVSDPASRL 274  
244 VVVQGLYKSM-----KFKPRILGMTKEEMMIFIEASSENPCSLYSVCYSPAERY 295  
275 VDEAKVYLLPYERPMQGPRTKPRKPLQGSSEVLA--VG-----CGWCSGDAIASI 323  
296 YKL-----CSPADLHKVGT--VTPNDIYIAGGOVPLKNTKTNHSTKSLQTAFT 344  
324 ERLDPMKGGTTWKVAPMGKRCGVAVLENLVYAVGG-----HDGQSYLNSYE 373  
345 RTVNCFYMPDAQONTWPK-TPLMFLVIRKPSLVCEGYIYAIGDSVGGELNRRTERVY 403  
374 R-----YDPMTNQSSVAPATCRISVGAARNGSLYAVGGD--GESCLDVEKID 424  
404 TEKDEWTVVSPPLPCAMQWSAAVVVHDCIYV-----TLNLMYCYFPRSDSVEMAMQ 456  
425 PKNEWAKYASMGTRRLGVSVVNGCIYAVGSGNPAPLNTVERIDPRVGMKEEVRPML 484  
QY 457 TSPSPASAAAFDCKIFYIGGLIAT 481  
DB 485 TRKRLGTAIVDGYIYAVGGRTTT 509

RESULT 4

T37322  
kelch-like protein kel-1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
R/Ohmachi, M.  
Submitted to the EMBL Data Library, February 1999  
A/Description: Kel-1, a novel Kelch-related gene in C.elegans, is expressed in pharyngea  
A/Reference number: Z21687

A:Accession: T37322  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-618 <OHM>  
A:Cross-references: EMBL:AB023424; PIDD:BAA82800.1  
R:Gajdathy, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z19209  
A:Accession: T20001  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-618 <WIL>  
A:Cross-references: EMBL:Z65902; PIDD:CAA93769.1; GSPDB:GM00020; CESP:C47D12.7  
A:Experimental source: clone C47D12  
C:GeneticS:  
A:Gene: kel-1; C47D12.7  
A:Map position: 2  
A:Introns: 512/2; 97/2; 201/2; 255/2; 489/3; 538/3; 565/1; 599/3  
C:Superfamily: POZ domain homology  
E:42-160/Domain: POZ domain homology <POZ>

Query Match	13.3%;	Score 438;	DB 2;	Length 618;
Best Local Similarity	24.2%;	Pred. No. 3.4e-26;		
Matches 129;	Conservative 83;	Mismatches 206;	Indels 114;	Gaps 18

```

Qy 19 EQLKLFEEQLFPIVL---IVEGT-----EPCHRVLAACSSYFPAFMMSG 63
Db 44 QNLNLRRCQGECLDVLLEAVPIQGISSEDMGSDVDHGHIMHRVILSSASSYFPAMTGG 103
Qy 64 LBSKQTHVHLRNVDATLOIITTAATGNLAMDSTVQLYETACPLEVEDVLQRCREY 123
Db 104 LRSTORIIPIKEVDVEVLSOLDIWMYTRMRIDISQNOFILATFASLQITCVNDAQARF 163
Qy 124 LIRKINAEVNCVRLSEFADLFSCCEELKOSAKRWEKHTAVVYHODAFMOLSHDLIDLISS 183
Db 164 MLEBLDMTNCYCGAAEFRRARAHACQLAHAAQLYTRGHFPEIIDNEBLLSDBDACELIOD 222
Qy 184 DNLNVEKETVREAAMLMEYNTESHQYLSLVLSQIRIDAL-SEVTRAWFQGLPNDK 242
Db 224 DRITVSEKVMQAVLNWVAHEPSPRRPFLAELMSNVHLPLFGDDYLFEKM-----NDE 278
Qy 243 -----SVVQGLYKSMWPKFKRMLCMTETEBMMIFLASBENPCLSVSCYSPQAE 299
Db 279 MITSDAACLANVIEBGM-----QLKITQDI-----STEP-----HYPIQE 315
Qy 294 ---KYVYKLSR-PADLHKVGVTPDNDIYIAGGVPLKPTKTNHKSKTQLQAFRTVC 349
Db 316 LNRKMFARAREBPESQH-----IMVGGQAPRA:TVND----- 348
Qy 350 FYWEPDAQNTWPFKRTMLPVRIKPLVCGEGYIYAGDSVGGELNRETVRYXTEDEW 405
Db 349 --LFPDPSQLMSSCASLPQRRCSGSMCNIGVYITGG--ENGQARVASVFPYPRDTW 404
Qy 410 TWVSPLEPCMAWQSAVVVVDICIVM-----TNTLMYC-YF--PRSDWEMMROTSRFA 462
Db 405 RSNQMSARSTHGITTCCQOVLYAVAGFGDTGLASAEYIDPHGNWFFLPFSMSTRSSV 466
Qy 463 SAAPFGDKIYITGGLHATNSGRLPBSGVQDSSV---TYEIIYVNCNEK 510
Db 465 GVAAFEEDIYAIAGF-----DGVSKQCLNLTVEIFPRARHKR 501

```

RESULT 5  
T3322  
Hypochemical protein W02G9.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 28-Jul-2000  
C:Accession: T3322  
R:Rolling, T.; Wohlmann, P.; Bauer, C.; O'Neal, D.  
A:Submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid W02G9.  
A:Reference number: Z21504  
A:Accession: T3322

A:Status: Preliminary; translated from GS/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-836 <ROH>  
A:Cross-references: EMBL:AF067946; PIDN:AACT1684.1; GSPDB:GN00023; CESP:W02G9.2  
A:Experimental source: strain Bristol N2; clone W02G9  
C:Genetics:  
A:Gene: CESP:W02G9.2  
A:Map position: 5  
A:Introns: 87/3; 114/1; 144/1; 181/1; 312/2; 352/3; 441/3; 776/2  
C:Superfamily: POZ domain homology  
F:201-302/Domain: POZ domain homology <POZ>

Query Match	12.6%;	Score 415;	DB 2;	Length 836;
Best Local Similarity	22.4%;	Pred. No. 3.4e-24;		
Matches 146;	Conservative 88;	Mismatches 223;	Indels 196;	Gaps 17;

```

0Y 10 NTEVAVSLBEOJLFYEQOULTFDIVLYEGEIEPCHKNVLTCSYFPAMFESGISESKO 69
Db 194 NOBOSKIMEQMRILRQTEBELCDVBLVAGSVIARHRIILAAASYFLOAMFTNGVENEKK 253
0Y 70 THVHLRVNDAATLOIITIVAYTGNLAMDSTVEQLYETACFLQVED----- 115
Db 254 LTIELQDIPRESVRIIVDYIYTDKIALTMNNVHQLIFATJVLQVSDVCAYHIELMTQY 313
0Y 116 -----VJRCGEVYLKIKMAENC 133
Db 314 DSPRQGCCELARLOIKFPGFOAFGLFVAPRPSRLSEPMYIVAVACQPLATMTITSHNC 373
0Y 134 VRLSPADLFSCBELKOSAKMVEHKKFPAVVHODAFMOLSHDLLIDLISDNLANKEET 193
Db 374 MSLVHPSDIYCTNLISIEDPASQPCIRKSPFNISIFPHLKSGLNRSPLANVSEQD 433
0Y 194 VREAAAML-----EYNTES--RSQYLSVLSQIRJIDAL----- 225
Db 434 VFETIYQLEFKTLRLHLGVNPLCEAVNSVYKNSOECREBIIISAMVLDAMTPSRKHPE 493
0Y 226 -----SEVTORANFOGLPBNDKSVVVGGLYKSMPEKFPRLGMTKEEMVIFIBASENP 279
Db 494 SNHENTSEYASMACPSLITASSS-----STSTPKSVAG-----AIFCAGRGKA 539
0Y 280 CSLYSV-CYSPQAEKYVKLCSPPADLHKVGTVTTPDNDIYIAGGOVPLKOTKTNHKS 338
Db 540 GGPSSVEAYDMRRNQIEVPDMMSQRHVG-VVSANGNLVLAIG-----HDGTA 588
0Y 339 KLQTAFRVNCFTVFDQAQNTWPFKTPMLFPRIRPSLVCCGCIYATIGDVSUGELNRT 398
Db 589 HLAATAEA-----FQPSIRQWKRIASMKTARGLIIVASIENVYIYAVG--LDOTTCYKT 639
0Y 399 VERYTEKDEWTWYSP-----PCAMOW----- 421
Db 640 VERDIDEBDESYTADMDVQRGGVAVIGRYVLPALIGNDGTSLSLETCERPDMPLDKKR 699
0Y 422 -----SAVVVHDCTIYV-----TLNLTMYCFPRSDSVTEMARQTSRSPASA 464
Db 700 IASMKRRRAGGVCVLDGVLVLAIGFDNNALETCERPDPADKVIITLDKNSSPRGGVGV 759
0Y 465 AAFPDKIFYIGSLHIATNTSGIRLBSGTVDGSSV--TVEIYVNNKMEQMAANI 515
Db 760 AALGGKVALIGG-H-----DGSDYLTNVECYPIARMKCPALET 797

```

RESULT 6  
T25373  
hypothetical protein T27E9.4 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 28-Jul-2000  
C:Accession: T25373  
R:Lloyd, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20024  
A:Accession: T25373  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA









Db 266 FTVVNNVLLYCLGFTL-----DGAFTCDVLA-----YDLITNEXSLMBEWH 308  
Qy 369 VRIKPLVCCGGTYIAGDSVGGELNR--RTVERDYTEKDEWTWVSLPCAMQWGAAYV 426  
Db 309 YRRMSACLVNGITYVGGVD--EENRLIGSEYVQPMGEH----- 348  
Qy 427 VHDCTY---VMTLNMV-----CYFP-----RSDSWEMANRQTSRSF 461  
Db 349 -HDAPYLAQANVETATVCYRNEELWIVGGYVDLPHFTISAVKULDRMRMMSEPLPFRSG 407  
Qy 462 AAAAAAGDITFYIGGHIATNSGIRLPSCGVDSSTVTLIVD--NKMEKMAANIPIAG 519  
Db 408 ATTIVVNNRLYCG-----GRHGGAYTNVYNYLDESRTERVGDYANVR 453  
Qy 520 YSDPCVRAVVISNSLCVPMRETHL-----NERAKYVTCYVDLEDRMSLRQHISERVLD 574  
Db 454 RNPSC-----CVYKAIYVGGNTNAVVKY-----NKKMQE----- 485  
Qy 575 LGRDFRCYVGL--YPSG 590  
Db 486 -----VGNISTYPAC 495

## RESULT 13

T24063  
hypothetical protein R09A8.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C/Accession: T24063  
R.Wilkinson, J.  
Submitted to the EMBL Data Library, November 1995  
A/Reference number: Z19836  
A/Accession: T24063  
A/Status: preliminary; translated from GB/EMBL/DDJ  
A/Molecule type: DNA  
A/Residues: 1-817 <MIL>  
A/Cross-references: EMBL:Z68009; PIDN:CAA92005.1; GSPDB:GN00028; CESP:R09A8.3  
A/Experimental source: clone R09A8  
A/Genetics:  
A/Gene: CESP:R09A8.3  
A/Map position: X  
A/Intons: 31/3; 77/3; 133/1; 198/2; 240/1; 282/3; 330/3; 440/1; 745/2; 779/1

Query Match 7.2%; Score 235.5; DB 2; Length 817;  
Best Local Similarity 16.9%; Pred. No. 4.2e-10;  
Matches 127; Conservative 123; Mismatches 269; Indels 233; Gaps 28;

Qy 17 LLEQLKLYEQQLFTDILVIVGTEPPCHKMLATCSTSYRAMF----- 60  
Db 80 LLDRLNTFRNRRELCDVLFVKEREIFVHKVLAASPALDMFNEENESSQDEKSPSS 139  
Qy 61 -----MSGLS-----ESKQTHVHNRVDAATLQIIITYATGNL 94  
Db 140 SEPTNGHTNGTNGNGEVAHENGAPVSAQKQKSTTFEYFQSDTFECFELVAFATYANL 199  
Qy 95 ANNDSTVEQLYETACFLOVEDVLRQCREYLKIKINENCVRLLSFADLSCEBELKOSAKR 154  
Db 200 EISRSKVAELKYKAVALQMTPIVIRACAAVLAENLQSLNCIGIRQANFNHDTLMKVKQ 259  
Qy 155 MNEHKTAAYYHQA-----FMQLSHDI 176  
Db 260 FIVDNDSIYNDSEKFTQLFVIOVRIIVPADEKVTNNANNGLAEMALFYQNNPHDR 319  
Qy 177 L--LILS-----SDNLNVEKEFTVREAAMLLEYNTESRQ----- 211  
Db 320 AEHSIDLLCTCKTHILYMEENHLDCIEMDRSSVSCDII-QDYKSGKDKRKQVAAKATA 378  
Qy 212 ---YLSSVLSQ-----IRI-----DALSEVTORAFQGLPNDK--SVVYQGLYKS 252  
Db 379 QEPITINSVQHRVNGAVPRLNLSRLPMNMAWSNESLESAGTDSDDQDTEALRLSTHQ 438  
Qy 253 MPEKFP-----RLGNTKEEMMITEASSENPCSLYS-----VCYSPQA 292

Db 439 AFOYVAVLVLYRCLVLSLQITDNEELL--RTYREPCVPDSQKALLSLRISCTGKQ 495  
Qy 293 EKYYKLCSPPADLHKVTVTPNDIYIAG-----QVPIKQTKTNHSTSKLQ 341  
Db 496 KPLECMSAPRC---SIGASFL-NGKITVCGGIDRECLRSVEBYVAVNGKRN--VALMK 549  
Qy 342 TAFRTNCFYV-----FDAQONTWPKTMLPVRIKPSLVCCGG 380  
Db 550 ABRGAFDCTIHGGKFAVAGSNGNNDLKSCEIYDPRKADVMAPLPIATPARCHNGCATIDN 609  
Qy 381 YYYALG---DSVGEELNRRYTERDYTEKDEWTWVSLP-----CAWQSAAVVND 429  
Db 610 YVYCIGSGFDQKVLDCERLDTTKGTBAEWEPAWMEHPRYQAGVCTWK-GLVVAAGG 668  
Qy 430 CIYVNTLNMVYCFPRSDSWEMANRQTSRSPASAAAGDKIFYIGGHIATNSGIRLPS 489  
Db 669 CWRWCMSVEAFEDKTNAMKQLPRLQARRGAVAVVRDITLYVGG-HDGTQS----- 721  
Qy 490 GTVDSSTVEIYDVNKNEMKMAANIPAKRYSDPCVRAVISNSL--CV-----PMRE 540  
Db 722 --LD---TVEILDSPSSQWRVGPILTTPRAN--THAVVTAGNVIFCIGFNGVKFLDT 772  
Qy 541 THLNERAKVTYQVLEDRMSLRQHISERV 572  
Db 773 IELLEN-----EOIGWRNMQHCPQQL 794

## RESULT 14

W21ZB6  
59k HindIII-C protein - vaccinia virus (strain WR)  
C/Species: vaccinia virus  
C/Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999  
C/Accession: F33348  
R.Kotwal, G.J.; Moss, B.  
Virology 167, 524-537, 1988  
A/Title: Analysis of a large cluster of nonessential genes deleted from a vaccinia virus  
A/Reference number: A94385; MUID:89073756; PMID:2849238  
A/Accession: F33348  
A/Molecule type: DNA  
A/Residues: 1-512 <KOT>  
A/Cross-references: GB:M2812; NID:g335691; PIDN:AAA69606.1; PID:g893315  
C/Superfamily: vaccinia virus 59k HindIII-C protein

Query Match 7.0%; Score 232; DB 1; Length 512;  
Best Local Similarity 21.5%; Pred. No. 3.9e-10;  
Matches 116; Conservative 85; Mismatches 240; Indels 98; Gaps 22;

Qy 33 IVLIVGTEPPCHKMLATCSTSYRAMFWSGLSES--KQTHVHNRVDAATLQIIITYAY 90  
Db 4 VIFSINGELIQVKEII-TASPY--NFKRIQDHLDKDEAILNGJNVHAFESLDYIR 59  
Qy 91 TGNLAMDSTVQLYETACFLOVEDVLRQCREYLKIKINENCVRLLSFADLSCEBELKQ 150  
Db 60 WKKINTINNVMILVAIIIDVPPVVDLCVTMTHNINSTCINMFNSKRYGKIKLYN 119  
Qy 151 SAKRWENHFTAVYHQAQFMQLSHDLIDILSSDNLNVEKEFTVREAAMLLEYNTES-- 208  
Db 120 AAMSEIINNITAVTSDPREGKLSKDELTITLSHEVNVNHBVTMILKWHKQPNVD 179  
Qy 209 -----RSQYLSVLSQIRIDALSEVTORAFQGLPNDKSVVYQGL-YKSNPKFK 258  
Db 180 IINIILPKKMTWTMENAISLGL-TISKSTKPTVBNIGKH--IYIKSDYISTTHYS 236  
Qy 259 PR-----LGMTKEEMMITEASSENPCSLY-----SSVCYSPQAQKVKYKLCSP 301  
Db 237 PRTEYWTIVGNTDRC--FYNNVNLHC-LYIIGMINNRHAYSVSRVDLRTKKMTVTN 292  
Qy 302 PADLHKVGTVPVPPNDIYIAG-QVPLKNTKTNNHK--TSKLQTAFTVNCFYWFDQAQN 358  
Db 293 MSLSKSEVSTCNDKQKLYIGGLEFSISGVAEYLKHGSK----- 333  
Qy 359 TWPKTMLPVRIKPSLVCCGGIYIAGDSVGGELNRRYERY--DTE--XDEWTM 411



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:15:27 ; Search time 14.6056 Seconds  
(without alignments)  
4965.877 Million cell updates/sec

**Title:** US-09-815-379-2

Sequence: 1 MSTQDERQINTEYAVSLLEQ.....STDGTEEFELDGEMVALLPV 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pcp.*
2: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/prodata/2/pubppaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pcp.*
6: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pcp.*
10: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/prodata/2/pubppaa/US10_NEW_PUB.pcp.*
12: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pcp.*
14: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3292	100.0	573	9	US-09-815-379-2	Sequence 2, Appl1
2	880	26.7	615	9	US-10-801-481-17	Sequence 17, Appl1
3	565.5	17.2	605	9	US-10-097-065-237	Sequence 28, Appl1
4	509	15.5	532	9	US-09-374-046A-78	Sequence 77, Appl1
5	455	13.8	538	10	US-09-764-864-1250	Sequence 1250, Appl1
6	374	11.4	287	10	US-09-864-761-42836	Sequence 42836, Appl1
7	340.5	10.3	307	10	US-09-764-864-1222	Sequence 1222, Appl1
8	336.5	10.2	317	10	US-09-925-300-1708	Sequence 1708, Appl1
9	327	9.9	236	9	US-10-106-698-5100	Sequence 5100, Appl1
10	233.5	7.1	524	10	US-09-764-864-1243	Sequence 1243, Appl1
11	231	7.0	231	9	US-10-106-698-6227	Sequence 6227, Appl1
12	228	6.9	156	10	US-09-925-300-1039	Sequence 1039, Appl1
13	205.5	6.2	113	9	US-10-811-059-313	Sequence 313, Appl1
14	202	6.1	379	9	US-10-106-698-5372	Sequence 5372, Appl1
15	199	6.0	168	10	US-09-764-864-1223	Sequence 1223, Appl1
16	199	6.0	185	10	US-09-925-302-624	Sequence 624, Appl1
17	195.5	5.9	408	9	US-10-014-137-2	Sequence 2, Appl1
18	192.5	5.8	109	10	US-09-107-058-7	Sequence 7, Appl1
19	192.5	5.8	109	10	US-09-761-117-7	Sequence 7, Appl1

20	187	5.7	114	9	US-10-211-059-319	Sequence 319, App
21	180.5	5.5	115	9	US-10-211-059-322	Sequence 322, App
22	168.5	5.1	108	9	US-10-211-059-316	Sequence 316, App
23	168.5	5.1	111	9	US-10-211-059-321	Sequence 31, App
24	167	5.1	527	9	US-10-211-059-3	Sequence 3, App1
25	165	5.0	171	10	US-09-764-864-1225	Sequence 1225, App
26	163.5	5.0	111	9	US-10-211-059-314	Sequence 314, App
27	163.5	5.0	180	10	US-09-764-864-1224	Sequence 1224, App
28	163	5.0	123	10	US-09-867-550-828	Sequence 828, App
29	162.5	4.9	105	10	US-09-764-864-1237	Sequence 1237, App
30	153.5	4.7	195	10	US-09-764-864-1180	Sequence 1180, App
31	153	4.6	117	10	US-09-864-864-145806	Sequence 45806, App
32	151	4.6	164	9	US-10-097-065-564	Sequence 664, App
33	151	4.6	165	9	US-10-097-065-556	Sequence 256, App
34	148	4.5	104	9	US-10-164-359-2	Sequence 2, App1
35	147.5	4.5	1259	9	US-10-260-715-8	Sequence 8, App1
36	144	4.4	107	9	US-10-164-359-8	Sequence 8, App1
37	143	4.3	114	9	US-10-211-059-318	Sequence 318, App
38	143	4.3	243	9	US-10-076-622-507	Sequence 507, App
39	143	4.3	108	9	US-10-007-805-507	Sequence 507, App
40	140	4.3	151	9	US-10-164-359-7	Sequence 7, App1
41	140	4.3	641	10	US-09-764-864-1231	Sequence 1231, App
42	138.5	4.2	151	9	US-10-108-605-61	Sequence 61, App1
43	138	4.2	158	10	US-09-764-864-1237	Sequence 1237, App
44	138	4.2	199	10	US-09-764-864-1228	Sequence 1228, App
45	136.5	4.1	354	10	US-09-731-872-416	Sequence 416, App

## ALIGNMENTS

```

RESULT 1
US-09-815-379-2
; Sequence 2, Application US/09815379
; Publication No. US20030073613A1
GENERAL INFORMATION:
APPLICANT: RASTELLI, LUCA
APPLICANT: GERRITSEN, MARY
TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING THE SAME
FILE REFERENCE: 10716/35
CURRENT APPLICATION NUMBER: US/09/815,379
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 60/191,134
PRIOR FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 623
TYPE: PRT
ORGANISM: Homo sapiens
US-09-815-379-2

```

	Query Match	100.0%;	Score 3292;	DB 9;	length 623;	
	Best Local Similarity	100.0%;	Pred. No. 1.8e-284;			
	Matches 623;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0	
Qy	1	MSIDDERQINTEYAVSLLEQLKLFYEQQLFTDVIIVIEGTEPCKHVIATCSTYFRAMF				60
Dh	1	MSIDDERQINTEYAVSLLEQLKLFYEQQLFTDVIIVIEGTEPCKHVIATCSTYFRAMF				60
Qy	61	MSGISESKQTHVHVRNDADATLQIIITAYTGNLAMDNSTEQLYETACFLQVEVLOKRC				120
Dh	61	MSGISESKQTHVHVRNDADATLQIIITAYTGNLAMDNSTEQLYETACFLQVEVLOKRC				120
Qy	121	REYLIIKIKINAMNCVRLSLFADLFSCBELKQSAKAMVEHKFTAVYHQDAFMQLSHDLLIDI				180
Dh	121	REYLIIKIKINAMNCVRLSLFADLFSCBELKQSAKAMVEHKFTAVYHQDAFMQLSHDLLIDI				180
Qy	181	LSSNNLNAVKEGETYREAAAMLLENVTNERSQYLSSVQLRIIDALSBVTQRAWFGCLPPN				240
Dh	181	LSSNNLNAVKEGETYREAAAMLLENVTNERSQYLSSVQLRIIDALSBVTQRAWFGCLPPN				240

QY 241 DKSVMVQGLYKSNPKFPRGLGNTKEEMMIFIEASSENPCSLYSVCYSPOAEKVVYKCS 300  
DB 241 DKSVMVQGLYKSNPKFPRGLGNTKEEMMIFIEASSENPCSLYSVCYSPOAEKVVYKCS 300  
QY 301 PPADLHKVGVVTPNDNDIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQONTM 360  
DB 301 PPADLHKVGVVTPNDNDIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQONTM 360  
QY 361 PPTPMPLFVRIKPSLVCCEGYIYAIGDSVGGELNRTVRYTDEKDEMTWSPPLCAMQ 420  
DB 361 PPTPMPLFVRIKPSLVCCEGYIYAIGDSVGGELNRTVRYTDEKDEMTWSPPLCAMQ 420  
QY 421 WSAVVVHDCIYVMTNLAMYCFPRSDSWEMAMROTSSPASAAGDKIFYIGLHIA 480  
DB 421 WSAVVVHDCIYVMTNLAMYCFPRSDSWEMAMROTSSPASAAGDKIFYIGLHIA 480  
QY 481 TNSGIRLPSTGVGSSVTVVEIYVNNKEMKMAANIPAKRYSDDPCRAVVISNLSLCPMRE 540  
DB 481 TNSGIRLPSTGVGSSVTVVEIYVNNKEMKMAANIPAKRYSDDPCRAVVISNLSLCPMRE 540  
QY 541 THINERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCVTKLYPSCLSESPMKPPT 600  
DB 541 THINERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCVTKLYPSCLSESPMKPPT 600  
QY 601 YLFSTDTGTEEFELDGEWVALPV 623  
DB 601 YLFSTDTGTEEFELDGEWVALPV 623

RESULT 2  
US-10-201-481-17  
Sequence 17, Application US/10201481  
Publication No. US20030119024A1  
GENERAL INFORMATION:  
APPLICANT: Linsley, Peter  
APPLICANT: Mao, Mao  
APPLICANT: Biezy, Matthew  
TITLE OF INVENTION: Genes and Proteins Associated with T-Cell Activation  
FILE REFERENCE: 9301-133-999  
CURRENT APPLICATION NUMBER: US/10/201,481  
CURRENT FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,968  
PRIOR FILING DATE: 2001-07-20  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 17  
LENGTH: 575  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-481-17

Query Match 26.7%; Score 880; DB 9; Length 575;  
Best Local Similarity 33.3%; Pred. No. 1e-69;  
Matches 198; Conservative 110; Mismatches 247; Indels 40; Gaps 10;

QY 13 YAVSLLEQLKLFYEQQLFTDIVIVE-GTEFPGHKWLATCSSYFRAMENSGSESKQTH 71  
DB 13 YAVSLLEQLKLFYEQQLFTDIVIVE-GTEFPGHKWLATCSSYFRAMENSGSESKQTH 71  
QY 5 HACSILKQLKTMVDEGLTIDIVEVDHCKTFSCHRNLAISPYFSMTSGLTSTQKE 64  
DB 5 HACSILKQLKTMVDEGLTIDIVEVDHCKTFSCHRNLAISPYFSMTSGLTSTQKE 64  
QY 72 VHLRNDATLQIITTYAYGNLAMDSTVEQLYETACPLQVEDVLTORCREYILKIKINE 131  
DB 72 VHLRNDATLQIITTYAYGNLAMDSTVEQLYETACPLQVEDVLTORCREYILKIKINE 131  
QY 65 VRIIVGEAESMDVLNAYATSRVILTEANVALFTAASIQIPISIDQCCKWIMISHDDP 124  
DB 65 VRIIVGEAESMDVLNAYATSRVILTEANVALFTAASIQIPISIDQCCKWIMISHDDP 124  
QY 132 NCYRLLSFADLFSCBELKOSAKRMVHKFTAAVYHQAQFMQISDLDLIDILSSNLNVEKE 191  
DB 132 NCYRLLSFADLFSCBELKOSAKRMVHKFTAAVYHQAQFMQISDLDLIDILSSNLNVEKE 191  
QY 125 NSIGVFIFADHYGHQELGDRSKKEYIRKFLCVTKGEFLLTYOQLSILSDSDLVNDR 184  
DB 125 NSIGVFIFADHYGHQELGDRSKKEYIRKFLCVTKGEFLLTYOQLSILSDSDLVNDR 184  
QY 192 ETVREAAMLLLEVYTESRSQYLSVLSQ-IRIDALSEVTORAWFOGLPPNDKSVVQGLY 250  
DB 192 ETVREAAMLLLEVYTESRSQYLSVLSQ-IRIDALSEVTORAWFOGLPPNDKSVVQGLY 250  
QY 185 EHYVESIIRMFHQBONREHVLPEIFAKCIRPLMBD---TIEKIPQFAQAIKSCV 240  
DB 185 EHYVESIIRMFHQBONREHVLPEIFAKCIRPLMBD---TIEKIPQFAQAIKSCV 240  
QY 251 KSNPKFFK---PLUGMTKEEMMIFIEASSENPCSLYSVCYSPOAEKVVYKCSPPADLHK 307  
DB 251 KSNPKFFK---PLUGMTKEEMMIFIEASSENPCSLYSVCYSPOAEKVVYKCSPPADLHK 307

DB 241 EKGPSNTNGCTQRLGNTASEMILCFDAHAKHSKQKQVPCLDIITGVVFKLCKPPNDLRE 300  
QY 308 VGVVTVTPNDNDIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQONTMFPKTPML 367  
DB 301 VGLIVSPNDNDIYIAGGYR-----SSSEVSIDHQAENDFMVYDSTNMLSKPSLL 351  
QY 368 FVRIKPSLVCCEGYIYAIGDSVGE--LNRTVRYTDEKDEMTWSPPLCAMQMSAAV 425  
DB 368 FVRIKPSLVCCEGYIYAIGDSVGE--LNRTVRYTDEKDEMTWSPPLCAMQMSAAV 425  
QY 352 RARIGCKLYVCCGKMAIAGRYEGDGRNLSKVSCEYDSRENCMTTVCAMPVAMEFHNAY 411  
DB 426 VVHDCIYVMTNLAMYCFPRSDSWEMAMROTSSPASAAGDKIFYIGLHIAITNSGI 485  
QY 412 EYKEKIYVLOGEFFLEFEPQKDYWGFLTPMTVPRIQGLAAVYKDSIYYIAG--TCGNQ 468  
DB 486 RLPSGTVDSSVTVVEIYVNNKEMKMAANIPAKRYSDDPCRAVVISNLSLCPMRETHL-- 543  
QY 469 RM-----FTVEYIDELNKTTRKQDFPCDQSNIPFLKVLQNKHLHFVRATQYV 519  
DB 544 -----NERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCVTKLYPSCLSE 593  
QY 520 EEHVFRTRKNSLQYQYDDIADQMKVYETPDR-LMDLGRHFCAVAKLYPCQLQK 573

RESULT 3  
US-10-097-065-237  
Sequence 237, Application US/10097065  
Publication No. US20030055236A1  
GENERAL INFORMATION:  
APPLICANT: Moore, Paul A. et al.  
TITLE OF INVENTION: 110 Human Secreted Proteins  
FILE REFERENCE: P2021P1  
CURRENT APPLICATION NUMBER: US/10/097,065  
CURRENT FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: PCT/US98/27059  
PRIOR FILING DATE: 1998-12-17  
PRIOR APPLICATION NUMBER: 60/070,923  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,007  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,057  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,006  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,369  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,367  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,368  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,169  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,053  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,064  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,054  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,008  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,365  
PRIOR FILING DATE: 1997-12-19  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 237  
LENGTH: 609  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-097-065-237

Query Match 17.2%; Score 565.5; DB 9; Length 609;  
Best Local Similarity 26.3%; Pred. No. 1.2e-41;  
Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

Best Local Similarity 25.6%; Pred. No. 1e-36;  
Matches 157; Conservative 95; Mismatches 234; Indels 128; Gaps 21

```
OY      36  IVEGTERPGCHKVYLATCSYFPRAMSGLSKQTHNHLRNVDATLQIIITTAATGNLA 95
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1  MVOERKIPARHVVLAASHFFNLMFTTWMLESKFEBVLKQABPDIIIEOLVAFYATIS 60
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      96  MNDSTYEOULYETACFLQVDVLDQRREVLIKKIAENQVNLISPADLFSCEELKQSAKM 155
      96  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61  VSNNNVQSLIDANAQOILEPVKMCVDLKEQVNASNCLGISVLAECGLDCEBELKATPDDF 120
      61  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      156  VEHKFTAVVYHODAFMQLSHDLIDLISDNNINVEKEETVREAAATLAEVETESRQVLS 215
      156  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121  IHHFTLEVYKTDDEFQLDVKRVTHLLNDLTLTVAEDQVYDAARVLTKEBERKQRPMPVD 180
      121  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      216  VLSQIRIDALS-----VTQRAFGQLRPNDKSVVYGLYSKM 253
      216  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      181  ILAKVFPRLSKNPLSKTVQAEPLIQDNPBCKLVVISGMRYHLSFEDREBLVDG----- 235
      181  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      254  PRFFKRLGMEKEMMIFIEASS-NPQSLYSVCYSPQAEKYKLCSPRADLHKGTVV 312
      254  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      236  -----PRPRKKGIDYRIALFGGSGQPSGRYPNPXDYS-----WTDIRCFPEKRDACV 284
      236  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      313  TPNDIYIAGGQ--VPLKNTKTNHNSKSLQTAFTVNCFWPDAQONTMF---PKTPM 366
      313  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      285  FMDNVVYILGGSGLPFIK-----MDG--YVNVDSMTSKLGPPTP- 323
      285  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      367  LFVRIPIPSLVCC--EGYIYALGDSVGGELNRRITVERDYDEKDEM-TWVSRLP--CAMQW 421
      367  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      324  -----RDSLAAQABSKITTSGGSVGNQA-LYLEECIDRTBESMHIKPSMLTQRCG--- 374
      324  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      422  SAAVVVHDCIYV-----MTLNMICYPPRSDSVWEMAKQTSRSPASAAAFGDK 470
      422  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      375  HGVMEVANGIYVCGGSGLSNNVSGRYLANSCEVVDPATETWTBELCPMIELARKNHGLVFPKDX 434
      375  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      471  IFYIGLHIAITNSGIRLPBGTVDSSVVEIYDVNKKEMKMAANIPAKRYSDPCVRAVVI 530
      471  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      435  IFAVGG-----QNGL-----GGLDNVEYYDIIKLEMKVNSPMPKGVTVKCAAVGSI 481
      435  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      531  SNLSLCFMREETHINERAKYVTVQYDELDRWSLROHISERVLMDIGSDPFRCTVSKLYPSC 590
      531  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      482  VVYLAGFQGVGRIGH-----IIQVTEITDKVNAVSAQV-----RAFPVT--KLPLNLC 525
      482  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      591  LEESEPWKPPYLF 604
      591  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      526  CR-----YLWS 531
      526  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-09-764-864-1250
US-09-764-864-1250
US-09-764-864-1250, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acid, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1250
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (263)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (273)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
```

```

/ LOCATION: (332)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (346)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (347)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (351)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (394)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1250

Query Match      13.8%; Score 455; DB 10; Length 538;
Best Local Similarity 24.7%; Pred. No. 6.7e-32;
Matches 129; Conservative 97; Mismatches 206; Indels 90; Gaps 13;

QY 10 NTEVAVSLLEQLKLYEQQLFTDVLIVEGTEFPCHKMVLATCSYPRAMFMSGLSESKQ 69
DB 84 NT-HAASILMSMSLRKSNLTLCVTLRAVEQKDPARHIVLAACSDYFCAMFTSELSKXK 142
QY 70 THVHLRNVDAATLQIIITVAYTGNLAMNDSTVEQLYETACFLOVEDVLQRCREYLJKIN 129
DB 143 PYVDIOQLTASTWEILLDFYETETVHTVENVOELLPACLLQKGVKQACCFLESQDL 202
QY 130 AENCVRLLSPADLFSCBELKQSAKRWVHKFTAVYHODAMQIASHDLIDILSSDNLNVE 189
DB 203 PSCNCLGRDAETHNCVDLMQAAEVFSQKHFPEVVOHEBELLISQGEVEKLKICDEIQVD 262
QY 190 KEETVREAAMLMEYNTESRSQYLSVLSQIRIDALSEVYQRAMFQGLPNDKSVVYQGL 249
DB 263 XEPEVPEAVIXWYKAKKERESLPILQYKMPILF----- 239
QY 250 YKSMPEFKPRKLCNTKEEMMIFLEASSENPCSLYSVCSY--POAEKRYKLCSPPADLHK 307
DB 300 ----PRYITD-----VIDAEPFIRCSLQ---CRDLVDEAKKFF--LKPBEIRSQM 339
QY 308 VGTVTVPPNDI-YIAGQVVLKNTKTNHSTSKLQTRAVNCFYFPAQONWPFKTPM 366
DB 340 QGPRTRXXLDKMYVSGG---FGSRRHTSMER-----YDPNIDQMSMLGDM 382
QY 367 LPRIKPSLVCCEGYIYAIGDSVSGELNRTYERYDTEKDEMTWSPPLCAMQWSAAVY 426
DB 383 QTRBEGGLVYXGVIYCLGGYDGLNLN--SVKIDPHGHNTVTTPMAYKXSGAGVAL 440
QY 427 VHDCLYVM-----TLNLMYCYFPRSDSVWEMAMRQTSRSPASAAAFGDKIFYIGLHI 479
DB 441 LNDHIYVVGFGPDGTAHLSSVEAVNIRDSWTVTSMTPRCYGVATVLRGLVAIAG--- 497
QY 480 ATNSGIRLPSGTVDGSSV--TVELIDVKNKEMKMANIPAKR 519
DB 498 -----YDGNLSLSIECYDPIIDSWEVVSMGTQR 527

RESULT 6
US-09-864-761-42836
/ Sequence 42836, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aeonica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04

```

```

/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/532,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 42836
/ LENGTH: 287
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE: MAP TO AL049870.1
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.96
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
/ OTHER INFORMATION: EST HUMAN HIT: U9560.1, EVALUE 3.00e-55
/ OTHER INFORMATION: SWISSPROT HIT: Q04652, EVALUE 3.00e-57
US-09-864-761-42836

Query Match      11.4%; Score 374; DB 10; Length 287;
Best Local Similarity 36.0%; Pred. No. 4.4e-25;
Matches 77; Conservative 48; Mismatches 89; Indels 0; Gaps 0;

QY 13 YAVSLLEQLKLYEQQLFTDVLIVEGTEFPCHKMVLATCSYPRAMFMSGLSESKQTHV 72
DB 17 HSEQLQGLNLRRQHEHLCDDILRVGDVKIHAHKVYLVSVSYFFKAMFGNISEKENSEV 76
QY 73 HURNVDAATLQIIITVAYTGNLAMNDSTVEQLYETACFLOVEDVLQRCREYLJKINMEN 132
DB 77 BRQCIDETRLQAIIVEAYATGVIFISODYVESLPPANILLQITLVKECCAFLESQDPCN 136
QY 133 CYRLISFADLFSCBELKQSAKRWVHKFTAVYHQDAFMQIASHDLIDILSSDNLNVEKE 192
DB 137 CGISRFATYVCRLYLAAATYICNFEAVCQTEBFEFLTHADDEIVSNDCLNVAEE 196
QY 193 TVREAAMLMEYNTESRSQYLSVLSQIRIDALS 226
DB 197 TVFYALBSWIKYDVQBRQKYLQGLNSVRLPILS 230

```



Db 212 PEEB 215

RESULT 10  
US-09-764-864-1243  
Sequence 1243, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

Prior Filing Date: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1243

LENGTH: 224

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (15)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (94)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (185)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (212)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-864-1243

Query Match

Best Local Similarity 7.1%; Score 233.5; DB 10; Length 224;

Matches 55; Conservative 43; Mismatches 77; Indels 7; Gaps 4;

Db

10 NREYVSL--LEDDKLFYEQLFTDVLIV--EGTE-FPCXKVLACSSFFRAMEGSL 65

45 NTHSVVLGGFDDRL--EGLLCDVTLVPGDDELFPVHRAMMASADVPKXMTGGMK 101

66 ESKQTHVHLNRYVADATLQIITVAYTGNLAMDSTVEQLYETACFLOVEDVLCRCREYLI 125

102 EQLDMKIKLHGKVGKGLKTIIDITVAKLSLNDNLDOTLEAASFQILPVDPCVPLI 161

Db

126 KKIINACVRLISFADLFSCGELKQSAKRVKFTAVYHQDAFMQLSHDLIDILSSDN 185

162 SGVSLDNCVEVGRIANTVNLIEVKKYNNFLLKNFPALLSTGEFLKLPFXLAFLVSSNS 221

Qy

186 LN 187

Db

222 LS 223

RESULT 11

US-10-106-698-6227

Sequence 6227, Application US/10106698

Publication No. US20030109690A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

FILE REFERENCE: PA005PI

CURRENT APPLICATION NUMBER: US/10/106,698

Prior Filing Date: 2002-03-27

Prior Application Number: PCT/US00/26524

Prior Filing Date: 2000-09-28

Prior Application Number: US 60/157,137

Prior Filing Date: 1999-09-29

Prior Application Number: US 60/163,280

Prior Filing Date: 1999-11-03

NUMBER OF SEQ ID NOS: 8564  
SOFTWARE: Patentin Ver. 3.0  
SEQ ID NO 6227  
LENGTH: 521  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (84)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: MISC FEATURE  
LOCATION: (93)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: MISC FEATURE  
LOCATION: (94)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-106-698-6227

Query Match

Best Local Similarity 7.0%; Score 231; DB 9; Length 521;

Matches 101; Conservative 90; Mismatches 214; Indels 190; Gaps 21;

Db

122 EYLKIKINANCVRLLSFADLFSCGELKQSAKRVKFTAVYHQDAFMQLSHDLIDL 181

2 DYLSRMDVTSCISYRNFAFCWDSRLINKVDAYIQEHLQISEEFKLPKPR-LKLEVM 60

182 SSDNLNVEKEFTVREAMLMLEYNTERSQVLSVLSQIR-----IDALS 226

61 LEDNVCILPSNGKLYRKVIWVQRXIWNGDSLXXLMEEVQTLVYADHKLIDGNLIDGQA 120

227 EV-----TORAWFQGLPNDKS----- 243

121 EVFGSDDDHIFVQKKPRENGHKQISSSTGCLSPNATVQSPKHEWKIVASEKTSNNT 180

244 ----VVQGLY-----KMPKFK-----PRGCM 263

181 YLCIAVLGIFCVIFLAGNSPQSSPTSTPKLSKLSFEMQDELIEKMSPMQYARSGL 240

264 TKEEM--MIFEASENPCSLSVSVCSPOAEKYYKCSF--PADLHKVTVTTPND 317

241 GRAENKGLIAGGYNREEC-LRTVECTNPHIDH-WFLAPKRTTRARQAVLM--GQ 295

318 IYIAGQVPLKNTKTNHSTKSLQTAFTVNGCFWFDQAQNTWPFKTPMLFVRIKPSLVC 377

296 LVVVGQ-----SNHSD-----DLSCGEYDSDIWDIMFVPELRTRCNAGVCA 339

378 CEGYIYALGSDVSGELNRTYERYDTEKDEMTWSPPLPCAMQNSAAVYVHDCIYVM--- 434

340 INKGLYIVGSDPYQKGLKNCDDVPVTKMTSCAPLNIHRRQSAVCELGGLYLTIGGA 399

435 ----TLNLMYCYFPRSDSVEMAMRQTSRSPASAAAFGDKFYIGLHATNSGIRLPSG 490

400 ESMNCLNVERYNPENNITLIPMVARRGAVANGLKFLVCCGF----- 446

491 TVDGSSV--TVEIYDVKNEWMAANIPAKRYSDPCVRAVVISNLCVFMRETHNERAK 548

447 --DGHAIISCEVMYDPTREMKMGMTSPRSN--AGIATVGNTI----- 487

Qy

549 YTYOYDLELDRLWSLRHISERVLMIDGRDRCCTYGLYPSCLTESPMKPPYTLF 603

488 YAVGGFD-----GNEFLNTV-EVYN--LESNEWSPTYKIF 519

Db

RESULT 12

US-09-925-300-1039

Sequence 1039, Application US/09925300

Patent No. US20020151681A1

GENERAL INFORMATION:

APPLICANT: Steve Ruben

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA101

CURRENT APPLICATION NUMBER: US/09/925,300



```
/ CURRENT FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05988
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ PRIOR FILING DATE: 1998-03-12
/ NUMBER OF SEQ ID NOS: 1890
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1039
/ LENGTH: 156
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-09-925-300-1039
```

```
Query Match      6.9%; Score 228; DB 10; Length 156;
Best Local Similarity 35.0%; Pred. No. 1.8e-12;
Matches 43; Conservative 27; Mismatches 53; Indels 0; Gaps 0;
```

```
QY 9 INTERAVALLEQLKLYEQLFLDIYLYEGTEFPCHKNVLAACSSYFFAMNSGLSEK 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 32 VSEDHGOIKLSTYQNFREQNVFDFKIKMDEIIPCHRCVLAACSDFFRAMEVNMKERD 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 QTHVLRVNDANTQIIITVATGNLAMDSTVEQLYACFLQVEDVLRQREYLKIKI 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 92 DGSVTTMLSSKAVKAFDPAAYTGKXTTDDNVEMFQLSSFLQVSFLSKACSDPLIKSI 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 NAE 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 152 NLE 154
```

```
RESULT 13
US-10-211-059-313
/ Sequence 313, Application US/10211059
/ Publication No. US20030100495A1
/ GENERAL INFORMATION:
/ APPLICANT: Zhang, Jian
/ TITLE OF INVENTION: HUMAN NAC-1 PROTEIN
/ FILE REFERENCE: PB0149
/ CURRENT APPLICATION NUMBER: US/10/211,059
/ CURRENT FILING DATE: 2002-08-02
/ PRIOR APPLICATION NUMBER: US 60/311,034
/ PRIOR FILING DATE: 2001-08-08
/ NUMBER OF SEQ ID NOS: 322
/ SOFTWARE: Aecmics Sequence Listing Engine
/ SEQ ID NO 313
/ LENGTH: 113
/ TYPE: PRF
/ ORGANISM: Artificial Sequence
/ OTHER INFORMATION: Consensus sequence of BTB motif
US-10-211-059-313
```

```
Query Match      6.2%; Score 205.5; DB 9; Length 113;
Best Local Similarity 38.1%; Pred. No. 1.2e-10;
Matches 43; Conservative 20; Mismatches 49; Indels 1; Gaps 1;
```

```
QY 16 SLLEQLKLYEQLFLDIYLYEGTEFPCHKNVLAACSSYFFAMNSGLSEKQTHVLR 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 SLQNLINERNGEFCDVTLVVGKKEFKAKHAKVLAACSPYFKLFGNFKES-SSITTD 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 NYDAATLQIIITVATGNLAMDSTVEQLYACFLQVEDVLRQREYLKIKI 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 DVSVEFALBERITGKLEITENVELELADKQIGSLVDLGEFLIKNI 113
```

```
RESULT 14
US-10-106-698-5372
/ Sequence 5372, Application US/10106698
/ Publication No. US20030109690A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
/ FILE REFERENCE: PA005PI
```

```
/ CURRENT APPLICATION NUMBER: US/10/106,698
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: PCT/US00/26524
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US 60/157,137
/ PRIOR FILING DATE: 1999-09-29
/ PRIOR APPLICATION NUMBER: US 60/163,280
/ PRIOR FILING DATE: 1999-11-03
/ NUMBER OF SEQ ID NOS: 8564
/ SOFTWARE: PatentIn Ver. 3.0
/ SEQ ID NO 5372
/ LENGTH: 379
/ TYPE: PRF
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (10)_
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5372
```

```
Query Match      6.1%; Score 202; DB 9; Length 379;
Best Local Similarity 23.9%; Pred. No. 1.4e-09;
Matches 92; Conservative 62; Mismatches 171; Indels 60; Gaps 15;
```

```
QY 201 WLEYNTERSOYLSVLSQIRIDALSE--VTQRAWFGGLPPND--KSVVQGL-YKSNP 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 WIKYDVEROKYLAQLNSVRLPLSLVKFLFRLYBANHLIRDRCKHLNBLAKYHFW 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 255 K-----FFKPLGKMTKEEMIFIEASSNPSCSYSSVCSQAERVTGLCPPADL 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 73 EHRLSHQVLMTRPCA---PKVLCAGVGKSGFLFACLSVEMYFPQNSWIGLAPLNTPR 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 HKVGTVPNDNDIYAGGOVPLKNTKTNHSTKSKQTAFRTVNCYWFPAQONTWPPKTP 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 YEGICVL-DQKVTYIGG-----IATVPRGVTTIKENSVEC--KNPDNTVTSLE 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 MLFVRIKSLVCEGYIYAIGDSVSGELNRTTVERDYDEKDEWTVSBLPCAMQMSAAV 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 180 KMSRSTLGVVLADELVALGG--YDGSYLQSVKEYIPIKIKMQFVAPMTTRSCFALA 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 VVHDCIYVM-----TILMLCYFPRSDSVEMAKRQSRASAAABODKIFYIGLHI 479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 VLDGMITVIAIGGYGPAHMSVERYPDSDKSWEMVAMADKRIFHGVGMGLFVVG-- 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 ATNSGIRLPSTGVDSVTVVEIYDVKEMKMAANIPAKRYSDPCRAVINSICV--- 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 --HNGVSHLS-----SIERYDPHONQWYCRMKERTG--VGAADVINYLYVVG 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 537 FMRETHLNERAKYVTQYDLLEDRW 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 342 HSGSSYLN-----TVQKYDPISDTW 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 15
US-09-764-864-1223
/ Sequence 1223, Application US/09764864
/ Patent No. US20020132753A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PT223
/ CURRENT APPLICATION NUMBER: US/09/764,864
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PAM or file wrapper
/ NUMBER OF SEQ ID NOS: 1792
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1223
/ LENGTH: 168
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-09-764-864-1223
```

```
Query Match      6.0%; Score 199; DB 10; Length 168;
```

Best Local Similarity 27.8%; Pred. No. 7.8e-10;  
Matches 42; Conservative 39; Mismatches 52; Indels 18; Gaps 2;

```

Qy 11 TEAVSLLEQLK-----FYEQQLFTDIVLIVEGTEFPCHKVLATCSSYPRAMF 60
    ||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 16 TEKKLAAREAKLLAGFMGMNNRRKOKTLCDVILWQERKIPAHRVVLAASHFFNLMF 75
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 61 MSGLSSEKQTHVHLRVNDAATLQIITVAYTGNLANNDSTVEQLYETACFLQVEDVLRQ 120
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 76 TTNMLSSKSPFVELKDAEPDIIEQLVEFAVTARISVNSNNVQSLDPAANGYQIEPVKWC 135
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 121 REYLIIKINAEKNCVRLISFADLFSCSELKOS 151
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 136 VDPLKEQVDASNC-----LGEAEKVDOS 158
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: July 14, 2003, 18:26:55  
Job time : 16.6056 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 6.47994 Seconds  
(without alignments)  
3987.653 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292  
Sequence: 1 MSTDQERQINTEYAVSLLEQ.....STDGIEPELDGEMVALPVPV 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	572	17.4	606	1 KRPI_RAT	Q9e330 ratu
2	569	17.3	1477	1 KRLC_DROME	Q04652 drosophila
3	565.5	17.2	609	1 KRLP_HUMAN	Q9y2m5 homo sapien
4	554	16.8	606	1 KRLP_HUMAN	Q06662 homo sapien
5	545.5	16.6	587	1 KHL2_HUMAN	Q9uh77 homo sapien
6	534	16.2	593	1 KHL2_HUMAN	Q95198 homo sapien
7	523	15.9	748	1 KHL1_HUMAN	Q9n164 homo sapien
8	516.5	15.7	589	1 ENCI_MOUSE	Q35709 mus musculu
9	516	15.7	751	1 KHL1_MOUSE	Q9j174 mus musculu
10	511.5	15.5	589	1 ENCI_HUMAN	Q14682 homo sapien
11	505.5	15.4	718	1 KHL4_HUMAN	Q9c0h6 homo sapien
12	501	15.2	755	1 KHL5_HUMAN	Q96997 homo sapien
13	492	14.9	624	1 KEAP_HUMAN	Q14145 homo sapien
14	474	14.4	632	1 YD54_HUMAN	Q9p2j3 homo sapien
15	473	14.4	624	1 KEAP_MOUSE	Q9z2x8 mus musculu
16	467	14.2	624	1 KEAP_RAT	P57790 ratu
17	455.5	13.8	584	1 IPP_HUMAN	Q9y573 homo sapien
18	455	13.8	639	1 YD05_HUMAN	Q9p2n7 homo sapien
19	445	13.5	652	1 YD84_HUMAN	Q9p293 homo sapien
20	443.5	13.5	539	1 Y466_HUMAN	Q9u1p4 homo sapien
21	418.5	12.7	584	1 IPP_MOUSE	P28515 mus musculu
22	399	12.1	597	1 GAN_HUMAN	Q9h2c0 homo sapien
23	345.5	10.5	465	1 Y795_HUMAN	Q94889 homo sapien
24	327	9.9	564	1 VAS5_VACCC	P21073 vaccinia vl
25	324	9.8	564	1 VAS5_VACCC	P24768 vaccinia vl
26	309.5	9.4	530	1 VC04_SPVKA	P32228 swinepox vl
27	287	8.7	588	1 CALI_HUMAN	Q13939 homo sapien
28	284	8.6	588	1 CALI_BOVIN	Q28068 bos taurus
29	273	8.3	515	1 VMT6_MYXVL	P22611 myxoma viru
30	251	7.6	623	1 Y711_HUMAN	Q94819 homo sapien
31	244.5	7.4	509	1 VMT9_MYXVL	P08073 myxoma viru
32	234.5	7.1	500	1 VC13_SPVKA	P32206 swinepox vl
33	232	7.0	512	1 VC02_VACCV	P17371 vaccinia vl

34	227	6.9	512	1 VC02_VACCC	P21037 vaccinia vl
35	226	6.9	579	1 YR47_CAEBL	Q09563 caenorthabdi
36	216	6.6	480	1 VF03_VACCV	P24357 vaccinia vl
37	208	6.3	480	1 VF03_VACCC	P21013 vaccinia vl
38	202.5	6.2	441	1 YD40_HUMAN	Q9p2j3 homo sapien
39	199	6.0	374	1 SPOF_HUMAN	C43791 homo sapien
40	195	5.9	918	1 SCRA_TIMPO	Q25390 timulus pol
41	190.5	5.8	522	1 BTB3_HUMAN	Q9y2f9 homo sapien
42	188.5	5.7	532	1 BTB3_MOUSE	F58545 mus musculu
43	183.5	5.6	518	1 YNV6_CAEBL	P34569 caenorthabdi
44	180.5	5.5	525	1 YNV2_HUMAN	C9b770 homo sapien
45	178.5	5.4	274	1 YKV2_CAEBL	F34324 caenorthabdi

## ALIGNMENTS

```

RESULT 1
KRPI_RAT
AC Q9ER30;
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DE Kelch-related protein 1 (Kel-like protein 23) (Sarcosin).
GN KRPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic;
RX MEDLINE=20180502; PubMed=10713668;
RA Spence H.J., Johnston I.P., Bwart K., Buchanan S.J., Fitzgerald U.,
RA Ozanne B.W.;
RT "Krlp1, a novel kelch related protein that is involved in pseudopod
RT elongation in transformed cells.";
RL Oncogene 19:1266-1276(2000).
CC - FUNCTION: REQUIRED FOR PSEUDOPOD ELONGATION IN TRANSFORMED CELLS.
CC - SUBCELLULAR LOCATION: PREDOMINANTLY CYTOPLASMIC BUT CAN CO-
CC LOCALIZE WITH F-ACTIN AT THE MEMBRANE RUFFLE-LIKE STRUCTURES AT
CC THE TIPS OF TRANSFORMATION-SPECIFIC PSEUDOPODIA.
CC - TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN SKELETAL MUSCLE. ALSO
CC FOUND IN HEART AND LUNG.
CC - SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC - SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ293948; CAC08185.1;
CC InterPro: IPR000210; BTB_POZ.
CC InterPro: IPR001798; Kelch.
CC Pfam: PF00651; BTB; 1.
CC Pfam: PF01344; Kelch; 5.
CC SMART: SM00225; BTB; 1.
CC PROSITE: PSS0097; BTB; 1.
CC Cytokeleton; Repeat.
CC DOMAIN
CC FT 33 100
CC FT REPEAT 346 398
CC FT REPEAT 399 447
CC FT REPEAT 448 495
CC FT REPEAT 497 542
CC FT REPEAT 544 599
CC FT SEQUENCE 606 AA; 68213 MW; 08DB54E9298DACE CRC64;
SQ
Query Match 17.4%; Score 572; DB 1; Length 606;

```

Best Local Similarity 27.0%; Pred. No. 2.2e-34;  
Matches 157; Conservative 95; Mismatches 236; Indels 94; Gaps 13;

[illegible]

RESULT 2			
KELC_DROME	STANDARD;	PRT;	1477 AA.
ID_KELC_DROME	Q04653; Q9Y0A2;		
AC	Q04652; Q04653; Q9Y0A2;		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Ring canal kelch protein [Contains: Kelch short protein].		
DE	KELC OR CG7210.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
NCBI	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RX	MEDLINE=93201592; PubMed=8453663;		
RA	Xue P., Cooley L.;		
RT	"Kelch encodes a component of intercellular bridges in Drosophila egg		
RL	Cell 72:681-693(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkelley;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adame M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Mananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		

RA Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayan A., An H.-U., Andrews-Plamroch C., Baldwin D.,  
RA Ballew R.M., Baay A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Bokova D., Bochan M.R., Bouck J., Brooksstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dublin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman I.J., Hernandez J.R., Houck J.,  
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitai M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., Meleod M.P., McPherson D.,  
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclob J.M.,  
RA Palazzolo M., Peltman G.S., Pan S., Plessard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Syrakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinlock G.M., Weissbach J.,  
RA Williams S.M., Wodgett T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
[3]  
RA CHARACTERIZATION.  
RC TISSUE=EMBRYO;  
RC MEDLINE=97236487; PubMed=9118811;  
RX Robinson D.N., Cooley L.,  
RA "Examination of the function of two kelch proteins generated by stop  
RT codon suppression";  
RL Development 124:1405-1417(1997).  
CC -1- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF  
CC CYTOPLASM BETWEEN CELLS. MAY BE INVOLVED IN THE REGULATION OF  
CC CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.  
CC BINDS ACTIN.  
CC -1- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING  
CC CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK  
CC EPITHELIA.  
CC -1- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE  
CC TESTIS, OVARECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND  
CC IMAGINAL DISKS. KELCH ORP1 IS THE PREDOMINANT PROTEIN AND IS ALSO  
CC EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF  
CC FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.  
CC -1- DEVELOPMENTAL STAGE: LARVAE, PUPAE AND ADULTS.  
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
CC -1- CAUTION: REF.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.  
CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
CC gene model prediction.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/annotation/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC  
CC EMBL, L08463; AAA53471.1; -  
DR EMBL, L08463; AAA53472.2; -  
CC EMBL, AE003657; AA053651.1; ALT\_SEQ.

DR HSSP; Q05516; 1CS3.  
DR FlyBase; Fggn0001301; kel.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR001798; Kelch.  
DR Pfam; PR00651; BTB; 1.  
DR Pfam; PR01344; Kelch; 6.  
DR PRINTS; PR00501; KELCHREPEAT.  
DR SMART; SM00225; BTB; 1.  
DR PROSITE; PSS0097; BTB; 1.  
DR Cytokeleton; Actin-binding; Selenium; Selenocysteine; Repeat.  
KW Cytokeleton; Actin-binding; Selenium; Selenocysteine; Repeat.  
FT CHAIN 1 1477  
FT DOMAIN 1 689  
FT REPEAT 157 223  
FT REPEAT 404 449  
FT REPEAT 450 496  
FT REPEAT 498 543  
FT REPEAT 545 592  
FT REPEAT 594 639  
FT REPEAT 641 687  
FT DOMAIN 18 28  
FT DOMAIN 29 36  
FT DOMAIN 78 83  
FT SE\_CYS 690 690  
FT CONFLICT 493 493  
FT CONFLICT 596 596  
FT CONFLICT 824 824  
FT CONFLICT 858 858  
FT CONFLICT 1083 1083  
FT CONFLICT 1086 1086  
SQ SEQUENCE 1477 AA; 160086 MW; 4851EAE9D9DBA47 CRC64;

Query Match Best Local Similarity 17.3%; Score 569; DB 1; Length 1477;

Matches 189; Conservative 104; Mismatches 249; Indels 98; Gaps 22;

QY 10 NTEYAVSLLEQLKLFYEQQLFTDIVLIVEGTEPCHRWYATGSSYFRAMFMSGLSESKQ 69  
DB 136 NQCHTARSPDANMEKQKQOLCVIIVADVEIHAHRMVAACSPYFAMFMS--PESRQ 194  
QY 70 THVHLNVDATLQIIITTAATGNLAMDSTVEQLYETAACFLOVEVQLQRCREYLKIKIN 129  
DB 195 ARITLQSVARALELLIDVYATVAVENEDNVQLLTAAQLQTLTVRACCCDQLDTOLD 254  
QY 130 AENCVALLSFADLFSCCELKQSAKRWKFTAVYHQAEMOASHDLIDILSSDULNTE 189  
DB 255 ASNCGLGIRFPADHACVELLNATYETIEQFNENVIQPDENLNSHQVSLINDKISVP 314  
QY 190 KEETVREAAMLWLEYNTESRQYLSVLSQIRIDALSE--VTQRAWFOGLPNDKSVVQ 247  
DB 315 NEBRVVECVIAMLRYDVPWREQFTSLMEHVRLPFLSKETITQRV-----DKELILE 366  
QY 248 G-----LYSMKPKFK-----PR--LGMTKEEMFTIASSENPCSLYSVCY 288  
DB 367 GNIVCKNLIIETALTYLLPFTKSAATVPKPKPMKPKM-----IILVIGQAFAKIRSEV 422  
QY 289 SPOAEKYVYLCSPPADLHKVGVTVTDNDIYINGGVPLINTKNSKTSKIQTARTYN 348  
DB 423 DLREKRYQAEMPNRRCRSGSLSLDCK-VYAVG-----FNSS-----LRYRTVD 467  
QY 349 CFYMPDAQONTWPKPTMFLVRIKPSLVCEGYIYVAGISVSGELNRRTERVYDEKXB 408  
DB 468 V---YDPATDQVANCNMEARSTLGVAVLNGCIVAGG--FDGTTGLSAAEYDKPTDI 522  
QY 409 YTMVBLPCAMQMSAAVYVHDCIYV-----TLNLYCYCFPPSDSVEMAKQTSR 459  
DB 523 MRPIASMTSRSSVGVVHGLIYVAGVDFTRQCLSSVERVNPDTDMVNVVAEMSSSR 582  
QY 460 SFASAAAFGDKIYIGGLHATNSGIRLPSGTVDGSSV--TVEIYVNKNEMGMANIRA 517  
DB 583 SGAGVGLNIIYAVGG-H-----DGPWNRSAVATCETNSMRSVADMSY 627  
QY 518 KRYSDCVAIVVINS--LCVPMRETHLERAKYVTVYQYDLRLDRSLRQHSERVLMWL 575

DB 628 CRRN-----AGVADHGLLYVGGDDGTSNLSAVEVYCPD--SDSMRLIP-----ALMTI 675  
QY 576 GRDPR--CTVGLKPSCLSE 593  
DB 676 GRSYAGVCKWIDK--DWCEE 693

RESULT 3 KHLX HUMAN

ID Q9Y2M5; Q9H457; STANDARD; PRT; 609 AA.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

OS Kelch-like protein X.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Yoshida K.; Sugano S.;

RT "Kelch motif containing protein."

RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 9-609 FROM N.A.

RA Frankland J.;

RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.

CC -1 SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.

CC -1 SIMILARITY: CONTAINS 6 KELCH REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib.ch](mailto:license@sib.ch)).

CC -----

DR EMBL; AB026190; BA77027.1; -

DR EMBL; AL109921; CAC10469.1; -

DR InterPro; IPR000210; BTB\_POZ.

DR InterPro; IPR001798; Kelch.

DR Pfam; PR00651; BTB; 1.

DR Pfam; PR01344; Kelch; 6.

DR SMART; SM00501; KELCHREPEAT.

DR PROSITE; PSS0097; BTB; 1.

KW Repeat.

FT DOMAIN 68 135  
FT REPEAT 319 365  
FT REPEAT 367 413  
FT REPEAT 414 460  
FT REPEAT 462 507  
FT REPEAT 509 554  
FT REPEAT 556 601  
FT CONFLICT 593 593  
SQ SEQUENCE 609 AA; 68083 MW; E7B2E3FE8341B422 CRC64;

Query Match Best Local Similarity 17.2%; Score 565.5; DB 1; Length 609;

Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

QY 9 INTEYAVSLLEQLKLFYEQQLFTDIVLIVEGTEPCHRWYATGSSYFRAMFMSGLSESK 68  
DB 46 ISDKPRQFLVEYINILRRKRELCDVVLVGAKKIYAHRIYLSACSPYFRAMFTGELASBR 105  
QY 69 QTHVHLNVDATLQIIITTAATGNLAMDSTVEQLYETAACFLOVEVQLQRCREYLKIKI 128  
DB 106 QTEVIRIDIDERAMELLIDFAYTSQITVEEGNVQTLPAACILQIAIEOACCEFLKROL 165

```

0Y 129 NANCYRLLSPPDLFSCCEELKQSGAKMVEKFAVYAHODAFMQLSHDLLDILSSDPLAV 188
Db 166 DSNCLGTAFAFDHDSCHRELLRADXKFTQHPQEBWSESEFLLPANQLDILISSDELAV
0Y 189 EKEETVREAMLMLEBYNTESRSGQYSSVLISQIRIDALS-----E 2277
Db 226 REEEQVFNAYVMWVKXSIQERRPQLPOVLOHVLRPLSPKFLVGTGSDPLIKSDECRD 285
0Y 228 VTQRAMFQGLPFPNDKSVVVQGLYKSNPKFKPPLGTMKEMLFIENSSNPSS---LYS 284
Db 286 LVDEAKNYLLRPOERP-LWQG-----PR-TREPKPIRCGVLFLAVGWM---CSGDAISS 3344
0Y 285 SVCYSPQAEKVYLLCSPPADLHKVGTVPVDPNDIYIAGS-----QVPLKNTKTHH-- 3344
Db 335 VERYPQPTNEMEMVAVMSMRKRCGVGSVL--DLLLVAGGHIDSSLYLNSVERYPKTNQWS 3533
0Y 335 ---SKTSKIQTA-----FRTVNCFWFMDAQONTWFPKPYMLFVRIRK 3737
Db 394 SDVAFTSTCRTEGVAVLGFLYAVAGDQGVSCSLNTVERYPDREKNKWTTRVASTRRLGV 4533
0Y 374 SLVCEGTYIYALGGDSVGGELNRRTERYDTEKDEKWTNWSPLCCANQMSAAVVVHHCITY 4533
Db 454 AVAVLGGFLYAVAGGSGTSPLN--TYERNPOENRHHTIAPKOTRRRHGLCAVYODTIVA 5111
0Y 434 M-----TLNLMYCYFPRSDSVEMAMQOTSRFASAAAFGDKIFIGLHATNSGIR 4866
Db 512 VCGRDDTTELSSAERYNPRTNQWSPVVAMTSRSGVGLAVVNNQLMVAVGF-----5622
0Y 487 LBSGTVDDGSSV--TYEIIYVNNKMEMAMNIPAKR 519
Db 563 -----DITTYLKTIEVEFDPDANVTMLYCGMNYR 591

```



```

OX NCB1_TaxID=9606;
RN [1]
RX MEDLINE=99326431; PubMed=10397770;
RA Soltyk-Espanola M.B., Rogers R.A., Jiang S., Kim T.A., Gaedigk R.,
RT "Characterization of Mayven, a novel actin-binding protein
RL Mol. Biol. Cell 10:2361-2375(1999).";
CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
CC THE BRAIN CELLS.
CC -1- SUBUNIT: BINDS ACTIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF059569; AAC67502.1; -.
CC DR HSSP; Q05516; 1CS3.
CC DR Genew; HGNC:6353; KHLH2.
CC DR MIM; 605774; -.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR001798; Ketch.
CC DR Pfam; PF00651; BTB; 1.
CC DR Pfam; PF01344; Ketch; 6.
CC DR PRINTS; PRO0501; KETCHREPEAT.
CC DR SMART; SM00225; BTB; 1.
CC DR PROSITE; PSS0097; BTB; 1.
CC DR Cytokeleton; Actin-binding; Repeat.
CC KM DOMAIN 56 123 BTB.
CC FT REPEAT 308 353 KETCH 1.
CC FT REPEAT 354 400 KETCH 2.
CC FT REPEAT 402 447 KETCH 3.
CC FT REPEAT 449 496 KETCH 4.
CC FT REPEAT 497 543 KETCH 5.
CC FT REPEAT 545 591 KETCH 6.
CC SQ SEQUENCE 593 AA; 66033 MW; 50479F44F359A81E CRC64;

Query Match 16.2%; Score 534; DB 1; Length 593;
Best Local Similarity 25.3%; Pred. No. 1.3e-31;
Matches 138; Conservative 100; Mismatches 199; Indels 108; Gaps 14;

QY 5 DEROINTE-----YAVSLLEQLKLFYEOOLFPTDIVLIVEGTEFPCHKVLA 51
DB 20 DSKDNDTEKHCPTVTPVPMHKKAFKVMNELR---SQNLCDVITVAEDMESHRRVLA 76
QY 52 CSSYFFPAMFSGISESKQTHVLRNDATLTQITTYATGNLANMDSTYEQLYTACFL 111
DB 77 CSFYFHAMFGEMSEBSAKRVRIKEVGMTLRLMLIDVYVAEIOVEENQVLLPAGLL 136
QY 112 QVEDVLCRCREYLIIKINENCRLISPADLFSCEELKQSAKMVHKFAVYVHQAQMG 171
DB 137 QLDQDVKTCCFEESQHPYNCIGIRFADMHACTDLNKAUTYAQHFADVLSEFPA 196
QY 172 LSHDLIDLSSDNLNVEKEETREANMLLEVNTESRSQYLSVLSQIRIDALSE--VT 229
DB 197 LGIEQYCSLISSKLTISSEKVFPAVIMVNDKVRQGFARMLMEHATLPLPREYLV 256
QY 230 QRAMPQGLPND--KSVVQGL-YKSMKFFPRILGTMTEEMIFTEASSENPCSLYS 285
DB 257 QRVEEBALVNSACKYKYLEAKMYLLP-----TEORILM----- 292
QY 286 VCVSPQAEKYKLCSPADLHKGTVTTPNDIYIAGGVPLKNTTNSKTSKQTARF 345
DB 293 -----KSVRTLRTPMNLKMLVVV-----GGQAP-----KAIR 321

```

```

QY 346 TVNCFYFPAQNTWPFKPTMLFVRIKPSLVCCGEIYVAGDSVGGELNRRTERYTE 405
DB 322 SVCC---YDFKEBWHQVLELPSRRRCRAGMVYAGLVFVAGG--FNGSLRVRTVSDYDV 376
QY 406 KPEWTVSPLPCANQMSAAVYVHDCIYV-----TLNLMTCYFPRDSWYEMARQTS 458
DB 377 KOQWTSVANMRDRRSTLGAIVANGLVAGGFDGSTGLSSVAYNFKSNEMFVAPMNR 436
QY 459 RSPASAAFGDKIFYGILHINTSGIRLPSGTVDGS-----VVEIYVUNKKEMWMAAN 514
DB 437 RSVGVGVGGLYVAGS-----YQASQYLSYTCVATINNEWYIAE 481
QY 515 IPAKR 519
DB 482 MSTR 486

RESULT 7
KHL1 HUMAN STANDARD; PRT; 748 AA.
AC Q9NR64; Q9NR65; Q9P238; Q9H4X4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ketch-like protein 1.
GN KHL1 OR KIAA1490.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RX MEDLINE=20347694; PubMed=1088605;
RA Koob M.D., Nemes J.P., Benzow K.A.;
RT "The SCA8 transcript is an antisense RNA to a brain-specific
RL transcript encoding a novel actin-binding protein (KHL1).";
RL Hum. Mol. Genet. 9:1543-1551(2000).
RN [2]
RX SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [3]
RX SEQUENCE OF 179-409 FROM N.A.
RA Kay M.;
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
CC THE BRAIN CELLS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF252283; AAF81719.1; -.
CC DR EMBL; AF252279; AAF81716.1; -.
CC DR EMBL; AB040923; BAA96014.1; ALT_INIT.
CC DR EMBL; AL353738; CAC16128.1; -.
CC DR Genew; HGNC:6352; KHLH1.
CC MIM; 605332; -.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR001798; Ketch.

```



```

DR Pfam; PF00651; BTB; 1.
DR Pfam; PF01344; Kelch; 6.
DR PRINTS; PR00501; KELCHREPEAT.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PSS00097; BTB; 1.
KM Cytoskeleton; Actin-binding; Repeat.
FT DOMAIN 43 88
FT REPEAT 212 279 BTB.
FT REPEAT 460 506 KELCH 1.
FT REPEAT 507 553 KELCH 2.
FT REPEAT 555 600 KELCH 3.
FT REPEAT 601 647 KELCH 4.
FT REPEAT 649 700 KELCH 5.
FT REPEAT 701 747 KELCH 6.
SQ SEQUENCE 748 AA; 82680 MW; C11C43D8282F9F9 CRC64;

Query Match 15.9%; Score 523; DB 1; Length 748;
Best Local Similarity 25.0%; Pred. No. 1,1e-30;
Matches 151; Conservative 101; Mismatches 237; Indels 114; Gaps 15;

QY 2 STODERQITVEYAVSLLEQQLFYEQQLFTDIYLYEGTEPFGKQVLTCSGYFRAMFM 61
DB 183 SSSSEFYQVHNAHQFRKKESTLYKQQLCDVLIYGNKRIHRLVLSVSDYFAMFM 242
QY 62 SGLSEKQTHVHLRNVDAAITLQIITYAYTGNLAMDSTVEQLYETACFLQVEDVLRGR 121
DB 243 SDVCEAKQSEIKMEGIDPNALMDLVQFAVTCLELEKEDTLENLLAAACLLQLPQVEVCC 302
QY 122 EYLKIKINENCYRRLSPADLFSCBELKQSAKMWENHKAFAVYHQAQFQSLDILIDL 181
DB 303 HFLMKLHPNSNCIGIAFADQCCILMKVAHSYTMENTWENIRNOEFLPLPEELHKL 362
QY 182 SSDNLVKEETVREAAMLLEYNTESRQYSSVLSQIRI-----DALSEVQRAMFOG 236
DB 363 ASDDVAVPBEETIYHLLMMWVKYDKMSRCNDLSMLAFIRLPLPQIADLENHLLFKK 422
QY 237 LEPNDKSVVQGLYKSMK-----FKPR----- 260
DB 423 DLEQQLILEAMKYLHLLPERRITMQSPRTKPRKSTVGTLYAVGMDNNKGATTIEKYDR 482
QY 261 -----LGMTKEEMAFIESSSENPCSLHSVYSQAEKVTYGLCSPPAD 304
DB 483 TNLMTQAGMNGRRLOFGVAVIDDKLFVIGRRGLKTLNVEECYNPKTKWTYL--PPMS 540
QY 305 LHKVGTVTTP-DNDIYIAGQVPLKTKTNHSTKSLQTFRTVNGCFYWFDAQNTWPFK 363
DB 541 TRHMGVTVLESPITAVGG-----HDSSTLNTVER-----WDPOSQMTFY 583
QY 364 TPLFVRIKPSLVCCEGYIYAGDSVGGELNRTYERVTYTEKDEWTNVSPLPCAMQWSA 423
DB 584 ASMSIARSTVGVAAALNGKLYSGGRD--GSSCLSSMEYYPHTNNKNNMCAPM--CKRRGV 640
QY 424 AVYVND-CIYVM-----TLNLMYCYFPSPDSVVENAKQTSRSPASAAAFGD 469
DB 641 GVAATCGFYAVAGHDAPASNHCSRLLDYERDPKTDITVMAVAPLSMPEDAVGVCILGD 700
QY 470 KIFYGGLHATNSGIRLPSGTWDGSSV--TVEIYVNNKIEGMAANIPAKRYSDCVRA 527
DB 701 RLTVAGG-----YDQGYTANTMESYPTQNTENMQASLNGR--AGACV-- 742
QY 528 VVI 530
DB 743 VVI 745

```

## RESULT 8

```

ENCL MOUSE STANDARD; PRT; 589 AA.
AC 035709;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-UTN-2002 (Rel. 41, Last annotation update)
DE Ectoderm-neural cortex-1 protein (ENC-1).

```

```

GN ENCL OR ENC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss albino; TISSUE=Brain;
RX MEDLINE=97252647; PubMed=9096139;
RA Hernandez M.-C., Andres-Barquin P.J., Martinez S., Bulfone A.,
RA Rubenstein J.L.R., Israel M.A.;
RT "ENC-1: a novel mammalian kelch-related gene specifically expressed in
RT the nervous system encodes an actin-binding protein."
RL J. Neurosci. 17:3038-3051(1997).
CC -1- FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF
CC NEURONAL PROCESS FORMATION AND IN DIFFERENTIATION OF NEURAL CREST
CC CELLS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. INTERACTS WITH THE ACTIN
CC CYTOSKELETON.
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN THE NERVOUS SYSTEM.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS HIGHLY DYNAMIC BUT MOSTLY
CC RESTRICTED TO THE NS. OUTSIDE THE NS, EXPRESSION IS DETECTED IN
CC THE ROSTRAL-MOST SOMITOMERE OF THE PRESOMITIC MESODERM, AT THE
CC TIMES CORRESPONDING TO THE EPITHELIALIZATION THAT PRECEDES SOMITE
CC FORMATION. FIRST DETECTED IN THE BRAIN AND SPINAL CHORD OF 12 PC
CC EMBRYOS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U65079; AAB64206.1; -.
CC GMD; MGJ:109610; Encl.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR001798; Kelch.
CC DR Pfam; PF00651; BTB; 1.
CC DR Pfam; PF01344; Kelch; 5.
CC DR SMART; SM00225; BTB; 1.
CC DR PROSITE; PSS00097; BTB; 1.
CC Actin-binding: Developmental protein; Cytoskeleton; Repeat.
FT DOMAIN 46 114
FT REPEAT 296 340
FT REPEAT 341 388
FT REPEAT 389 444
FT REPEAT 446 492
FT REPEAT 494 538
FT REPEAT 539 585
FT REPEAT 589 640
SQ SEQUENCE 589 AA; 66085 MW; 12B62354D508B6A2 CRC64;

Query Match 15.7%; Score 516.5; DB 1; Length 589;
Best Local Similarity 25.5%; Pred. No. 2.5e-30;
Matches 155; Conservative 94; Mismatches 237; Indels 122; Gaps 17;

QY 11 TEYAVSLLEQLFYEQQLFTDIYLYEGTEPFGKQVLTCSGYFRAMFMSGLSEKQT 70
DB 26 SSVADSVLTHNLNLKQORLFTDVLHAGNRTPCHRAVLAACSRFEAMFSGLKESQDS 85
QY 71 HYLNR-VDAATLQIITYAYTGNLAMDSTVEQLYETKCFYQVEDVLRGREYLYKKIN 129
DB 86 EVNFDNSIHPEVLELLDLVAYSSRYVINENASLLEAGDMLEFDIIRDACAFLEKNIH 145
QY 130 AENCYRLSPADLFSCBELKQSAKMWENHKAFAVYHQAQFQSLDILIDLISDNLNLYE 189
DB 146 PNLGCMLLSDAHQCTKLYELSWMCNSNFTIKSBEFLDLPQDMVVQGLSSBELFTE 205
QY 190 KEETVREAAMLLEYNTESRQYSSVLSQIRIDALSEVQRAMFOGCLPNDKSVVQGL 249

```

Db 206 DERLYESANMWSIDYDKKRYCYLPELLQTVRLALLPAIYIM-----ENVAMEEL 255  
 QY 250 YKSMPEKFKPRLGWTKE--EMMTFIEASSENPCSLVSVCSPO-----A 232  
 Db 256 ITQOKR-----SKEIVEAIRCKKIKIIONGVSTSLCARPKGTGHALFLAGGOTMC 307  
 QY 293 EKVY-----KLGSPPADL--HKVGTVPVTPDDIYIAGOVPLKNTKTHSKTSKLQTA 343  
 Db 308 DKLVLVDQKAKEIIPKADIPSPREFSACALGCKVYITGG-----RSGENGVSKD----- 357  
 QY 344 FRTVNCYWM-FDAQONTWPKPTPLFRIKPSLVCCGGIYIYAGDS-----VGGELN 335  
 Db 358 -----VWVYDTLHEEWSKAAPLVARFGHGAELKHCILYVGGHTAATGCLPASPSS 410  
 QY 396 RRTVERDTEKDEMTWSPPLPCAMQMSAAV-----VHDCIYVMTLIMYCY 442  
 Db 411 LKQVEQDPTTKNTKNTWVAPLRBGVSNAVSAKLKLPAGGTSVSHD-----KLPKVQCY 465  
 QY 443 FPRSDSVWEMAMQTSRSPASAAFGDKIFYIGLHIAATNSGIRLPSGTVDGSSVTEIY 502  
 Db 466 DGCENRVSVPATCPQPMRYTAAVLGNQIFIMG-----GDTEFSAGSAVKF 511  
 QY 503 DVAKNEKMAANIIPARKYDPCRAVVISNSLCYPRKETHLNRKATYVQ-----YD 555  
 Db 512 NSSTYQWTKVGDVYAKRMS---CHAVASGNKLYV-----GGYFGIQRCKTIDCYD 559  
 QY 556 LEIDRWS 562  
 Db 560 PTLIDVWN 566

RESULT 9  
 KHL1 MOUSE STANDARD; PRT; 751 AA.

AC 09J174;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Kelch-like protein 1.  
 GN KHL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
 NC NCB1 TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20347694; PubMed=10888605;  
 RA Koob M.D., Nemes J.P., Benzow K.A.;  
 RT "The SCAR transcript is an antisense RNA to a brain-specific  
 transcript encoding a novel actin-binding protein (KHL1).";  
 RL Hum. Mol. Genet. 9:1543-1551(2000).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF  
 THE BRAIN CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF252281; AAF81717.1;  
 DR MGD; MG1:2136335; K1H11.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001798; Kelch.  
 DR Pfam; PF00651; BTB\_1.  
 DR Pfam; PF01344; Kelch; 6.  
 DR PRINTS; PR00501; KELCHREPEAT.

DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 KW Cytoskeleton; Actin-binding; Repeat.  
 FT DOMAIN 45 90  
 FT REPEAT 215 282  
 FT REPEAT 463 509  
 FT REPEAT 510 556  
 FT REPEAT 558 603  
 FT REPEAT 604 650  
 FT REPEAT 652 703  
 FT REPEAT 704 750  
 SQ SEQUENCE 751 AA; 82932 MW; 6E846190CAC6B4C3 CRC64;

Query Match 15.7%; Score 516; DB 1; Length 751;  
 Best local similarity 24.9%; Pred. No. 3,7e-30;  
 Matches 150; Conservative 101; Mismatches 238; Indels 114; Gaps 15;

QY 2 STDERQINTEXAVSLLEQLKPYEQQLFTDYLIVLVEGTPEFCHKVTLATCSSYFRAMFM 61  
 Db 186 SSSSEFYQAVRHAEQGFRRKENVYLKQOOLCDVILLIGNKKIPAHRLVLSVSDYFAAMFT 245  
 QY 62 SGLSESKQTHVHLRNVDAITLQIITVATGULAMNDSTVEQLYETACLOVEDVLCQR 121  
 Db 246 SDVCEAKQKEITKMEGIDPNALMDLVQFATGCLCELEKEDITENLLAAACLDLPQVEVCC 305  
 QY 122 EYLKIKINENCVRILSFADLFCSEELKQSAKRWYEHKFTAVYHODAFMQLSHDLIDIL 181  
 Db 306 HFLMKLHPSNCLGIRAFADAGCIELMKVASYMENTMEVIRNGEFLFLPAEBELKHL 365  
 QY 182 SSDNLNVEKEFTVREAAIMLIEYNTESRSQYLSVLSQIRI-----DALEVTORAMFG 236  
 Db 366 ASDDVAVPDEETIFHALMMVVKYDMQRRCSDSLMLAFYRLDLPQIILADLENHALFYK 425  
 QY 237 LEPNDKSVVVGGLYSMPK-----FFKPR----- 260  
 Db 426 DLBCQKLLEAMKTHLLPERRTLMOSPTPKPKSVTGILYAVGMDNKGATTEIKYDIR 485  
 QY 261 -----LGWTKEMMIFIEASSENPCSLVSVCSPOAEKVKYLCSPAD 304  
 Db 486 TNLMTIQAGMNRRLQFGVAVIDDKLFVIGRGDGLKTLTAVTECYNPKTKTWTVL--PMS 543  
 QY 305 LHKVGTVTP-DNDIYIAGOVPLKNTKTHSKTSKLOTAFTVNCFCYFPAQONTWPK 363  
 Db 544 TRHGGIGTVLGGPIYAVGG-----HDGWSYLTNVER-----WDPOSQQWTVV 586  
 QY 364 TMLLFRIRKPSLVCCGGIYIATGDSVSGELNRRYTERVDTKDEMTWSPPLPCAMQMSA 423  
 Db 587 ASMSIARSTVGVAALNGKLYSVGRD--GSSCLSSMEYTDPTTKMSKCPM--CKKRGV 643  
 QY 424 AVVVDH-CIYV-----TLNLMCYFPRSDSVWEMAMQTSRSPASAAAFGD 469  
 Db 644 GVAICDGFPIYAVGGHAPASNNHCSRLDVEYEPRTDITWVAPLSPMRDAVGCCLDD 703  
 QY 470 KIFYIGLHIAATNSGIRLPSGTVDGSSV--TYEIVDNKNEKMAANIIPARKYDPCVRA 527  
 Db 704 RLYAVGG-----YDQGYLTNTWESYDPTQNTWOMASLNGR--AGACV-- 745  
 QY 528 VVI 530  
 Db 746 VVI 748  
 RESULT 10  
 ENCI1 HUMAN STANDARD; PRT; 589 AA.  
 AC O14682; Q9UPG9; O75464;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ectoderm-neural cortex-1 protein (ENCI-1) (p53-induced protein 10)  
 DE (Nuclear matrix protein NRP/B).  
 GN ENCI OR P1010 OR NRPB.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon cancer;  
 RA MEDLINE=97449378; PubMed=9305847;  
 RX Polyak K., Xia Y., Zweiler J.L., Kinzler K.W., Vogelstein B.;  
 RT "A model for p53-induced apoptosis.";  
 RL Nature 389:300-305 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98350113; PubMed=9683534;  
 RX Hernandez M.-C., Andres-Barquin P.J., Holt I., Israel M.A.;  
 RT "Cloning of human ENC-1 and evaluation of its expression and  
 regulation in nervous system tumors.";  
 RL Exp. Cell Res. 242:470-477 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hippocampus, and Fetal brain;  
 RA MEDLINE=98234394; PubMed=9566959;  
 RX Kim T.-A., Lim J., Ota S., Raja S., Rogers R., Rivnay B., Avraham H.,  
 RA Avraham S.;  
 RT "NRP/B, a novel nuclear matrix protein, associates with p110(RB) and  
 is involved in neuronal differentiation.";  
 RL J. Cell Biol. 141:553-566 (1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Struhsberg R.;  
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF  
 NEURONAL PROCESS FORMATION AND IN DIFFERENTIATION OF NEURAL CREST  
 CELLS. MAY BE DOWN-REGULATED IN NEUROBLASTOMA TUMORS.  
 CC -1- SUBUNIT: BINDS TO RAI. HYPOPHOSPHORYLATED RAI ASSOCIATES WITH ENCI  
 DURING NEURONAL DIFFERENTIATION, WHILE HYPERPHOSPHORYLATED RAI  
 ASSOCIATES WITH ENCI IN NONDIFFERENTIATING CELLS.  
 CC -1- TISSUE SPECIFICITY: DETECTED IN FETAL BRAIN TISSUE. MODERATE  
 EXPRESSION IN FETAL HEART, LUNG AND KIDNEY. HIGHLY EXPRESSED IN  
 ADULT BRAIN, PARTICULARLY HIGH IN THE HIPPOCAMPUS AND  
 AMYGDALA, AND SPINAL CHORD. DETECTABLE IN ADULT PANCREAS.  
 CC -1- DEVELOPMENTAL STAGE: DRAMATICALLY UPREGULATED UPON NEURONAL  
 DIFFERENTIATION.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL, AF010314; AAC39532.1; -;  
 DR EMBL, AF005381; AAC64498.1; -;  
 DR EMBL, AF059611; AAC26109.1; -;  
 DR EMBL, BC000418; AAH00418.1; -;  
 DR Genew; HGNC:3345; ENCI.  
 DR MIM; 605173; -;  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001798; Kelch.  
 DR Pfam; PF00651; BTB\_1.  
 DR Pfam; PF01344; Kelch; 5.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 KW Actin-binding; Developmental protein; Cytoskeleton; Repeat;  
 KW Phosphorylation.  
 FT DOMAIN 46 114 BTB.  
 FT REPEAT 296 340 KELCH 1.  
 FT REPEAT 341 388 KELCH 2.

FT REPEAT 389 444 KELCH 3.  
 FT REPEAT 446 492 KELCH 4.  
 FT REPEAT 494 538 KELCH 5.  
 FT REPEAT 539 585 KELCH 6.  
 FT CONFLICT 112 130 INEBNESLEAGDMLBFO -> HQLEGCRSLGSLVTC  
 FT CONFLICT 237 238 MSFK (IN REF. 1).  
 FT CONFLICT 402 402 RL -> TR (IN REF. 1).  
 FT CONFLICT 427 427 C -> S (IN REF. 2).  
 FT CONFLICT 430 438 V -> A (IN REF. 1).  
 FT CONFLICT 484 589 IREGVSMA -> RRRRYNCAO (IN REF. 1).  
 FT YTPAAVLNGQITMGDTERFSKSAKPSRYQNTKVDV  
 FT TAKMSCHAVASGNKLVVGGYFGIQKCTLDCDPTLW  
 FT NSITTVPSYLIPTAVSTWKLPDS -> IHSQSCPGGTQD  
 FT FLMGVIONFSACFCI (IN REF. 1).  
 SQ SEQUENCE 589 AA; 66129 MW; DB003AIDFA65BAA0 CRC64;  
 Query Match 15.5%; Score 511.5; DB 1; Length 589;  
 Best Local Similarity 25.4%; Pred. No. 5.7e-30;  
 Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;  
 QY 11 TEVAVSLLEQLKLFEEQLFDIVIVLEGTEPCCHQMTATSSYFRAMFMSGLSRKOT 70  
 DB 26 SSYADSVLTHNLMLKQRLFDVLHAGNRTPCRAVALAACSRIFBAFSGGLKESQDS 85  
 QY 71 HVHLRN-VDAATLQIITVAYTGNLAMDSTVEOLYETACFLQVEDVLOCRREYLKIKIN 129  
 DB 86 EVNPDNSIHPEVLELLDLYAASSRVIIINEENBSLLEAGDMLBFODIRACAEFLEKMLH 145  
 QY 130 AENCYRILSPADLFSCBELKOSAKMVEHKFVAVYHODAFMQLSHDLIDISSDNLNVE 189  
 DB 146 PTNCGMLLSDAHOCTKLYELSWMCNLSNFTIKRNBFDLPDPDVQQLSBELEFE 205  
 QY 190 KSEITREAAAMLLEYNTESRSQYLSVLSQIRIDLSVTQRAWFGQLPNDKSVVQGL 249  
 DB 206 DRLVYESAINWISTDLKKRYCYLPBELQTVLALIPALYILM-----ENVAMEEL 255  
 QY 250 YKSMKFFKPRLGMTKE--EMMIFTEASSENDCSLYSVCSPQ-----A 292  
 DB 256 ITRKRR-----SKEIYEAIRCKLIKLDGVVTSICAPRRTGHALFLGGQTFC 307  
 QY 293 EKVY-----KLCSPADI--HKVGYVTPNDIYIAGQVPLKNTKTNHSTSLQTA 343  
 DB 308 DKLIVYDOKAKIIPKADIPSPKKEFSACALGCKYVITGG--GSENGVSKD----- 357  
 QY 344 FRTVACFW-FAQGNTPMPEKTPMLFVRIKPSLVCEGVIYAIGDS-----VGGEIN 395  
 DB 358 -----VAVYDTLHEBSKAAPMLVARFGHSAELKCLVYVGGHTAATGCLPASPSVS 410  
 QY 396 RRTVERYDTEKDEMTMVSPLPCAMQMSAAV-----VHDCIYVNTLMLMYCY 442  
 DB 411 LKQVEHIDPTIKMTMVAPLREGVSNAAVVSAKLTFAPRGTSVSHD-----KLPRVQCY 465  
 QY 443 PRSSDSEWEMARQTSRSFASAAAFGDKI FYIGSLHIATNSGIRLPSTGTVDSSTVEIY 502  
 DB 466 DDCENRWTVPARCPOPMWRYTAAVAGNOIFIMG-----GDTEFSACSAKYKF 511  
 QY 503 DYNKKEWMAANIPAKRYSDPCVRAVVISNSICVFMRETHLNEBAKYVYV-----YD 555  
 DB 512 NSETVQMTKVGVTAKRWS--CHAVASGNKLYV-----GGYFGIQRCKTLDCYD 559  
 QY 556 ILELDKMS 562  
 DB 560 PTLDDVNN 566  
 RESULT 11  
 ID KHL4 HUMAN STANDARD; PRT; 718 AA.  
 AC 09COH6; 09Y3J5;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Kelch-like protein 4.

<p>GN KHLH4 OR KIAA1687.</p> <p>OS Homo sapiens (Human).</p> <p>OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>OX NCBI_TaxID=9606;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).</p> <p>RX MEDLINE=21295031; PubMed=11401425;</p> <p>RA Braybrook C., Warty G., Howell G., Anasoun A.,</p> <p>RA Moore G.E., Rose M.T., Stanier P.,</p> <p>RT "Identification and characterization of KHLH4, a novel human homologue of the Drosophila katch gene that maps within the X-linked cleft palate and ankyloglossia (CPX) critical region.";</p> <p>RL Genomics 72:128-136(2001).</p> <p>RN [2]</p> <p>RP SEQUENCE FROM N.A. (ISOFORM 1).</p> <p>RC TISSUE=Brain;</p> <p>RA MEDLINE=21082932; PubMed=11214970;</p> <p>RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;</p> <p>RT "Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";</p> <p>RL DNA Res. 7:347-355(2000).</p> <p>RN [3]</p> <p>RP SEQUENCE OF 133-718 FROM N.A. (ISOFORM 1).</p> <p>RA Isegai T., Oea T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,</p> <p>RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,</p> <p>RA Megatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,</p> <p>RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,</p> <p>RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,</p> <p>RA Nimomura K., Iwayanagi T.;</p> <p>RT "NDO human cDNA sequencing project.";</p> <p>RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.</p> <p>RN [4]</p> <p>RP SEQUENCE OF 150-718 FROM N.A. (ISOFORM 1).</p> <p>RL Bird C.;</p> <p>CC Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.</p> <p>CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/KHLH4c; are produced by alternative splicing.</p> <p>CC -1- TISSUE SPECIFICITY: Expressed in adult fibroblasts and in a range of fetal tissues including tongue, palate, and mandible.</p> <p>CC -1- SIMILARITY: CONTAINS 1 BTF/POZ DOMAIN.</p> <p>CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.</p> <p>CC -----</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.ebi.ac.uk/announcements">http://www.ebi.ac.uk/announcements</a> or send an email to <a href="mailto:license@ebi.ac.uk">license@ebi.ac.uk</a>).</p> <p>CC -----</p> <p>DR EMBL; AF284765; AAK9441.1; -</p> <p>DR EMBL; AF284766; AAK9442.1; -</p> <p>DR EMBL; AB051474; BAB21778.1; ALT_INIT.</p> <p>DR EMBL; AK022715; BAB14198.1; ALT_INIT.</p> <p>DR EMBL; AL035424; CAB39994.1; -</p> <p>DR Genew; HGNC:5355; KHLH4.</p> <p>DR MIM; 300348; -</p> <p>DR InterPro; IPR000210; BTF_POZ.</p> <p>DR InterPro; IPR001798; Kelch.</p> <p>DR Pfam; PF00651; BTF; 1.</p> <p>DR Pfam; PF01344; Kelch; 6</p> <p>DR PRINTS; PR00501; KELCHREPEAT.</p> <p>DR SMART; SM00225; BTF; 1.</p> <p>DR PROSITE; PS50097; BTF; 1.</p> <p>KW Cyoskeleton; Actin-binding; Repeat; Alternative splicing.</p> <p>FT DOMAIN 182 249</p> <p>FT REPEAT 430 476</p> <p>FT REPEAT 477 523</p> <p>FT REPEAT 525 570</p> <p>FT REPEAT 571 617</p> <p>FT KELCH 1</p> <p>FT KELCH 2</p> <p>FT KELCH 3</p> <p>FT KELCH 4</p>	<p>OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>OX NCBI_TaxID=9606;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).</p> <p>RX MEDLINE=21295031; PubMed=11401425;</p> <p>RA Braybrook C., Warty G., Howell G., Anasoun A.,</p> <p>RA Moore G.E., Rose M.T., Stanier P.,</p> <p>RT "Identification and characterization of KHLH4, a novel human homologue of the Drosophila katch gene that maps within the X-linked cleft palate and ankyloglossia (CPX) critical region.";</p> <p>RL Genomics 72:128-136(2001).</p> <p>RN [2]</p> <p>RP SEQUENCE FROM N.A. (ISOFORM 1).</p> <p>RC TISSUE=Brain;</p> <p>RA MEDLINE=21082932; PubMed=11214970;</p> <p>RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;</p> <p>RT "Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";</p> <p>RL DNA Res. 7:347-355(2000).</p> <p>RN [3]</p> <p>RP SEQUENCE OF 133-718 FROM N.A. (ISOFORM 1).</p> <p>RA Isegai T., Oea T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,</p> <p>RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,</p> <p>RA Megatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,</p> <p>RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,</p> <p>RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,</p> <p>RA Nimomura K., Iwayanagi T.;</p> <p>RT "NDO human cDNA sequencing project.";</p> <p>RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.</p> <p>RN [4]</p> <p>RP SEQUENCE OF 150-718 FROM N.A. (ISOFORM 1).</p> <p>RL Bird C.;</p> <p>CC Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.</p> <p>CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/KHLH4c; are produced by alternative splicing.</p> <p>CC -1- TISSUE SPECIFICITY: Expressed in adult fibroblasts and in a range of fetal tissues including tongue, palate, and mandible.</p> <p>CC -1- SIMILARITY: CONTAINS 1 BTF/POZ DOMAIN.</p> <p>CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.</p> <p>CC -----</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.ebi.ac.uk/announcements">http://www.ebi.ac.uk/announcements</a> or send an email to <a href="mailto:license@ebi.ac.uk">license@ebi.ac.uk</a>).</p> <p>CC -----</p> <p>DR EMBL; AF284765; AAK9441.1; -</p> <p>DR EMBL; AF284766; AAK9442.1; -</p> <p>DR EMBL; AB051474; BAB21778.1; ALT_INIT.</p> <p>DR EMBL; AK022715; BAB14198.1; ALT_INIT.</p> <p>DR EMBL; AL035424; CAB39994.1; -</p> <p>DR Genew; HGNC:5355; KHLH4.</p> <p>DR MIM; 300348; -</p> <p>DR InterPro; IPR000210; BTF_POZ.</p> <p>DR InterPro; IPR001798; Kelch.</p> <p>DR Pfam; PF00651; BTF; 1.</p> <p>DR Pfam; PF01344; Kelch; 6</p> <p>DR PRINTS; PR00501; KELCHREPEAT.</p> <p>DR SMART; SM00225; BTF; 1.</p> <p>DR PROSITE; PS50097; BTF; 1.</p> <p>KW Cyoskeleton; Actin-binding; Repeat; Alternative splicing.</p> <p>FT DOMAIN 182 249</p> <p>FT REPEAT 430 476</p> <p>FT REPEAT 477 523</p> <p>FT REPEAT 525 570</p> <p>FT REPEAT 571 617</p> <p>FT KELCH 1</p> <p>FT KELCH 2</p> <p>FT KELCH 3</p> <p>FT KELCH 4</p>
---	--

FT	REPEAT	619	670	KELCH 5.
FT	REPEAT	671	717	KELCH 6.
FT	VARSPLIC	700	718	EVPVYNGAGACVVYVKLP -> SMOELLQNFTYTKLMT
FT	SEQUENCE	718 AA;	80245 MW;	LGH (IN ISOFORM 2).
SO				D73855AAD638150A.CRC64;
	Query Match	15.4%;	Score 505.5;	DB 1; Length 718;
	Best Local Similarity	24.6%;	Pred. No. 2e-29;	
	Matches 148;	Conservative 103;	Mismatches 239;	Indels 111; Gaps 17
Qy	2	STODEFOINT-EYAVSLLEQLKLPFEOQLFPDILVLYEGTFEPFHKVLATGSSYFPAMF	60	
Db	152	ATSEEQFVFNHAAQOTLRKMEYKLEKQOLDVLLIGHLRIPARHLYLVSAVDYFAMF	211	
Qy	61	MSGLSSKQTHVHLNRADATQIITITAYTGNLAMDSTVEQLYEPACELOYEDVLCRC	120	
Db	212	TNDVLEAKKEBEVRMEGVDPNALNSLVQYATYGVQLQKEDTIESLAAACLLDTQYDVC	2711	
Qy	121	REYLIKKIENACVBLSPADLPSCSELKQSAKMEVHKFTAYVYHODAFMQLSHDLLIDI	1800	
Db	272	SNFLIKQHPNSCLGRISFGDAQGCTELLANHKYKMEHFLEVIKNOEFLLLPANEISKL	3311	
Qy	181	LSSDNLNVEKERTVEEAMLMLEXYTERSQYSSYSQIRI-----DALSEYTGAMFQ	2255	
Db	332	LCSDDDINVDDEETIFPHALQWGVGHVQNRQEGELMGLSYLRPLPLPQLADLETSSMFT	3911	
Qy	236	G-----LPPNDKSVV-----VQGLY-----KSMPEF--	2566	
Db	392	GDLEQCKLMEAMKXHLHPERRSMWQSPRTKPRKSTGALYAVAGMDAMKGTITIEKYL	451	
Qy	257	-----FPRIGMTKEEMITIBASSENPCGLSSVSYSPQAEKYKLCSPRA	303	
Db	452	RTNSMLHIGTNMGRLOFGVAVINDMKLVVGGRGGLKTLITVECFNP-VGKIVTV-NPMM	509	
Qy	304	DLHKVET-VVTPDNDIYIAGGVPLKMTKTNHSTKSLQTFARTVNFVWFMDAQONTWP	362	
Db	510	STHRHGLGVALTEGMYAVGG-----HDMGSLYNTYER-----WDEEGQMN	552	
Qy	363	KTPMLFVRIKPSLVCCGGYIYIAGDSVGEYNRRYERYDTEKENTWVSPLPCAMQWS	422	
Db	553	VASMSPTPRSTGVVALNNKLYAIGRD--GSSCLKSMSEYEPDHTNKWSLCAPIWSKRGV	610	
Qy	423	AAVYVHDCIYV-----TLNLWC--YEPKSDVEMAMRQTSRSPASAAAFD	469	
Db	611	GVATYVGFYLVGGHDAPASNHCSSLSPCVERYPKGSWSTVAPLSPDVAVVCPLGD	670	
Qy	470	KIFYIGLHIAITNSGIRLPSGTVDGSSV--TVEIYDVNKMEKMAANIPAKRYSDPCVRA	527	
Db	671	KLYVVG-----YDGHYTLNLTVESYDAQRNEMKEVYVNIQR-AGACVVV	714	
Qy	528	V 528		
Db	715	V 715		



FT REPEAT 327 372 KETCH 1.  
 FT REPEAT 373 423 KETCH 2.  
 FT REPEAT 424 470 KETCH 3.  
 FT REPEAT 471 517 KETCH 4.  
 FT REPEAT 518 564 KETCH 5.  
 FT REPEAT 565 611 KETCH 6.  
 SQ SEQUENCE 624 AA; 69665 MW; 70528F3BF43B6C90 CRC64;

Query Match 14.9%; Score 492; DB 1; Length 624;  
 Best Local Similarity 24.7%; Pred. No. 1,7e-28;  
 Matches 149; Conservative 112; Mismatches 227; Indels 116; Gaps 20;

10 NTEYAVSLLEQLKLFYEQQLFTDILVLE-----GTEPCHKVLAATCSSYFRAMFMSGL 64  
 59 HTKQAFGINMELRL--SQQL-CDVTLQVKYQDAPAFQAFMAHKVLAASSSVFRAFMTNGL 115  
 65 SESEKQTHVLRANDATLQIITATYATGNLAMDSTVEQLYERACFLQVDDVLCRCREYL 124  
 116 REGGMEVVSIEGHPKMERLIEFAATYASISMEKCVLHMNGAVMQLDSSVRAQSDFL 175  
 125 IKKINAMNCVRLSPADLFSCCELKQSAKRMVEHKFTAVYHODAFMQLSHDLIDILSSD 184  
 176 VOQLDSDMAIGANFAEQICVELHQARAYITMHGEGVAKQSEFFVLSHCQLVTLISRD 235  
 185 NLNVEKEETVREAAAMLLEVENTESRQYLSVLSQIRIDALSEVTORAMFQGLPPN---- 240  
 236 DLNVRCESEVFHACINMVKYDCQRRFYVALLRVAVRCHSLT-----PVPFLQW 283  
 241 --DKSVVVGGLYSMEKFFKPRILGKTKEMMITEASSENPGS-----LYSSVCYSPQ 291  
 284 QLOKCEILOSDSRC-----KDYLVKIFEEELTLH-KPQVWPCRAPKVGRILYTAGGYFR 337  
 292 AEKVYKLCSP-----ADLH-----KVGTVTPDNDIYIAGOVPLKATKTNHSTKSK 339  
 338 SLSTYLAHVPNSGTWLRADLOVPRSGLAGCV--GGLYAVAGR--NNSPGNTDSSA 392  
 340 LQTAFTVNCVFWDAQONTWPKPTMLFVRIKPSLVCCGTYIYAGDSVGGELNRRYV 399  
 393 L-----DC--YNPMTNQMSPCAPWSVPRNRIGVVIDIGHIYAVAGSH--GCIHNSV 440  
 400 ERYDTEKDEKWTMSPLPCAMQMSAAVVDCTIVM-----TLNLMYCFRPSDSWEM 452  
 441 ERERPRDEWHLVAFMLTRRIGVAVLNLRLYAVGGFDGTNLSAECYPPRNEMRM 500  
 453 AMQTSRSPASAAFGDKIFYIGL-----HIATNSGIR 486  
 501 TANNTIRSGAGVCLNCTIYAGGYGQDQLNVERYDVETETWTVAAPKAKHSALGIT 560  
 487 LPSGTV-----DGSV--TVEITYVNKEMKXANIPAKRYS-----DPCVRAVVIS 531  
 561 VHQGRIVYLGVDGHTFLDSVECCYDPTDTWSEVTMTSGRSGVAVTWMEPCRKQIDQ 620  
 532 NSLC 535  
 621 NCTC 624

Db

RESULT 14  
 YDS4\_HUMAN STANDARD; PRT; 632 AA.  
 AC Q9P2J3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein KIAA1354 (Fragment).  
 GN KIAA1354.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;

RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequence of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:65-73(2000).  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB037775; BAA92592.1; -  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001798; Ketch.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF01344; Ketch; 6.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 DR Hypothetical protein; Repeat.

FT NON TER 1 1  
 FT DOMAIN 1 134 BTB.  
 FT REPEAT 315 362 KETCH 1.  
 FT REPEAT 363 414 KETCH 2.  
 FT REPEAT 415 461 KETCH 3.  
 FT REPEAT 463 508 KETCH 4.  
 FT REPEAT 510 560 KETCH 5.  
 FT REPEAT 561 610 KETCH 6.  
 SQ SEQUENCE 632 AA; 71146 MW; ED8E7003432FD73E CRC64;

Query Match 14.4%; Score 474; DB 1; Length 632;  
 Best Local Similarity 25.6%; Pred. No. 3.5e-27;  
 Matches 163; Conservative 105; Mismatches 259; Indels 110; Gaps 22;

10 NTEYAVSL--LBOQLKLFYEQQLFTDILVLE--GTEPCHKVLAATCSSYFRAMFMSGLS 65  
 45 NTHSSVVLGGFQOLRL--EGLLCDVTLVPGGDELFPVHRMAMASASYFRAMFTGKMK 101  
 66 ESKQTHVLRANDATLQIITATYATGNLAMDSTVEQLYERACFLQVDDVLCRCREYL 125  
 102 EDDLMCKLHGVNKKGLKIIDIFITAKSLAMDNLDTLEAASFLQILPVLDFCVFI 161  
 126 KIKINAMNCVRLSPADLFSCCELKQSAKRMVEHKFTAVYHODAFMQLSHDLIDILSSD 185  
 162 SGVSLDNCVEVGRINANTYNTLIEVDKYVNNFILKNPPLALSTGEFLKLPPERLAFVLSNS 221  
 186 LNVEKEETVREAAAMLLEVENTESRQYLSVLSQIRIDALSEVTORAMFQGLPPNDKSVV 245  
 222 LKHCTELELFKAACRWLLE--DPRMDYAAKMKNR-----FPLMPODLIN 268  
 246 VOGL-----YKMPKFKPRL-----GMTKEEMK 269  
 269 VQTVDFMRRTDNCVNLLEASNYQMP--YMGVPMQSDRAIRSDTHLVTLGGVLRQQLV 327  
 270 ITEASSENPCSLYSVCYSPOAEKVVKLCSPADLHKVGVTVTPDNDIYIAGGOVPLKN 329  
 328 VKEKLK-----MYDERAQEWRSIAPADAPRYQGIYVI--GNFLYVVGQ----- 370  
 330 TITNNSKTSKQOTAPRTVNCVFWPDAQONTWPKPTMLFVRIRKPSLVCCGTYIYAGDS 389  
 371 --SNYD--TKGTYAVDTV--FRPDRYNNKMQVAVSLNEKRTFFHLISALGLHLYAAGGS 423  
 390 VGGELNRRTRVEYDTEKDEKWTMSPLPCAMQMSAAVVDCTIVM-----TLNLMYCY 442  
 424 AAGEL--ATVECYNRRAMNSYVAKMSBPHYAGAGVYGLMYISGITHDFQDLKMF 481  
 443 FPRSDSWEMAMQTSRSPASAAFGDKIFYIGLHIAITNSGIRLPSGTVDGSSV--TVEI 501

Db

Db 482 DBPTDKMKQKAPMTYVRLGLHCMTYGVGDKLYVIGGNHF-----RGTSYDVVLSCEY 532  
Qy 502 YVNNKEMWMAANIPAKRYSDDPCVRAVVISNSLCVFMRETHINERAKYVTOYDELDRM 561  
Db 533 YSPTLDQWTPPIAAM-LRGSD--VGVAVPEKKIYVGVGISMNRCNVEIVQKIDPKDM 589  
Qy 562 SLRQHSERVLWDLGDFRCYVQKLYPSCLSESPMKR 598  
Db 590 HKVFLDPE---SLGGIRACTLTIVFPF---EENPGSP 619  
RESULT 15  
KEAP MOUSE  
ID KEAP MOUSE STANDARD; PRT; 624 AA.  
AC Q922X8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Kelch-like ECH-associated protein 1 (Cytosolic inhibitor of Nrf2).  
GN KEAP1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN 11  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=99104163; PubMed=9887101;  
RA Itoh K., Wakabayashi N., Katoh Y., Ishii T., Igarashi K., Engel J.D.,  
RA Yamamoto M.;  
RT "Keap1 represses nuclear activation of antioxidant responsive elements  
RL by Nrf2 through binding to the amino-terminal Neh2 domain.";  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Lung;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamada I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Maetuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glasl C., King B., Kochiwa H.,  
RA Kuenl P., Lewis S., Matsuo Y., Nikaido I., Pasole G., Quackenbush J.,  
RA Schriml L.M., Staahl J.F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barch G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.U., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guefnicich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakemoto N.,  
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: RETAINS NFE2L2/NRF2 IN THE CYTOSOL THUS RESULTING IN THE  
CC SUPPRESSION OF ITS TRANSCRIPTIONAL ACTIVITY AND THE REPRESSION OF  
CC ANTIOXIDANT RESPONSE ELEMENT-MEDIATED DETOXIFYING ENZYME GENE  
CC EXPRESSION.  
CC -1- SUBUNIT: INTERACTS WITH THE N-TERMINAL REGULATORY DOMAIN OF  
CC NFE2L2/NRF2.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EMBL: AB020063; BAA34639.1; -  
CC EMBL: AK004738; BAB33519.1; -  
CC MGD: MGI:1858732; Keap1.  
CC InterPro: IPR000210; BTB\_POZ.  
CC InterPro: IPR001798; Kelch.  
CC Pfam: PF00651; BTB; 1.  
CC Pfam: PF01344; Kelch; 6.  
CC PRINTS: PRO0501; KELCHREPEAT.  
CC SMART: SM00225; BTB; 1.  
CC DR PROSITE: PS50097; BTB; 1.  
CC KW Transcription regulation; Repeat.  
CC FT DOMAIN 77 149  
FT REPEAT 327 372 KELCH 1.  
FT REPEAT 373 423 KELCH 2.  
FT REPEAT 424 470 KELCH 3.  
FT REPEAT 471 517 KELCH 4.  
FT REPEAT 519 564 KELCH 5.  
FT REPEAT 565 611 KELCH 6.  
SQ SEQUENCE 624 AA; 69552 MW; 4645DB0122FB5F54 CRC64;  
Query Match 14.4%; Score 473; DB 1; Length 624;  
Best Local Similarity 23.7%; Pred. No. 4.1e-27;  
Matches 145; Conservative 111; Mismatches 223; Indels 134; Gaps 19;  
Qy 10 NTEYAVSLLEOLKLFYEQOLFDTIVLYE-----GTEPCKHMYLATGCSYFRAMFMGL 64  
Db 59 HTKQAFGWNEDRL--SQOL-CDVTLQVYKEDIPAAQPMANHKVLAASSPVFKAFKFTGL 115  
Qy 65 SESKOTVHLNRVDAATLQIIITVAVTGNLANMNSTVEQLYETACFLQVEDYLCRCREYL 124  
Db 116 REQGEVVSISGIIHPKWERLIEPAVTASISGKCVLHWNMGAMVQIDSVRACSPFL 175  
Qy 125 IKKINAEKCVLLSRADLFSCBELKQSAKRVKFTVYHODAMQSLHLLIDILSD 184  
Db 176 VOQLDPSNAIGANPAEQIGCTELHQARREYIYMHFGVAQBEFFNLSHCOLATLISRD 235  
Qy 185 NINVEKETVREAAMLMEVNTESROYLSVLSQIRIDALSEVYQGFAMFQGLPNDXSV 244  
Db 236 DLNVCSESEVFRACTIDWKVDCPORRFYQALNRAVCHALT----- 277  
Qy 245 VVOGLYKSMKPFKFRLL-----GMTKEEM-IFLEASBNPCS-----L 282  
Db 278 -----PRLQVQLKCEILDADNCQYLVQIFQELTLHKPTCAVPCAPKVRGLI 328  
Qy 283 YSSVYCSQAQEKVYLCSP-----ADLH-----KVGTVTPNDYIYIAGGVPLKXT 330  
Db 329 YTAGGYFPQSLSYLEAVNPSPSGMWLRDLQVPPRSLAGCVV--GGLLYAVGGR---NNS 383  
Qy 331 KTNHSTKSLQTAFTVACFWFDQAQNTWPKTPMLVRIKPSLVCCSGYIYATGDSV 390  
Db 384 PDGNTDSSAL-----DC--YNPMTQWSPCASMSVPRNRIGVIGDGHLYAAGSSH- 432  
Qy 391 GGEILNRRTVERDYDEKDEMTWSPPLPCAMQWAAVVDCTIVM-----TLNLMCYF 443  
Db 433 -GCIHSSVEREPRDMDHLVAPMLTRIGVAVLRLLYAVAGPFTRLNSAEQY 491  
Qy 444 PRSDSWEMAKQTSRSPASAAAFGDKFYIGGL----- 477  
Db 492 PERNEMWMTITPMNTRISGAGVCLNLCIYAAAGYGGDQQLNSVERVDVETETWEVAPMR 551  
Qy 478 HIATNSGIRLPQGV-----DGSSV--TVEIYDNKEMWMAANIPAKRYS-----D 522  
Db 552 HHRSLAGITVHOGKLYVLGYDGHFTLDSVECDPDSPTWSEVTRMTSGRSGVAVTWME 611  
Qy 523 PCVRAVVISNSLC 535  
Db 612 PERKQIDQNCNC 624  
Search completed: July 14, 2003, 18:16:04  
Job time : 9.47994 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:07:56 ; Search time 28.9026 Seconds  
(without alignments)  
4441.379 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292  
Sequence: 1 MSTDQERQINTEYAVSLRQ.....STDGTEEFELDEMYALPEV 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_protein:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp Unclassified:\*
- 15: sp\_rv1rus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2732	82.7	511	4 Q9P239	Q9P239 homo sapien
2	2480	75.6	469	11 Q8RIW9	Q8RIW9 mus musculu
3	1886	57.3	348	4 Q9UFW7	Q9UFW7 homo sapien
4	1670	50.7	335	4 Q9Y382	Q9Y382 homo sapien
5	1664.5	50.6	351	4 Q9P097	Q9P097 homo sapien
6	778.5	23.6	526	4 Q9GJ15	Q9GJ15 homo sapien
7	638.5	19.4	684	4 Q8WZ29	Q8WZ29 homo sapien
8	631.5	19.2	674	4 Q96MP6	Q96MP6 homo sapien
9	630.5	19.2	684	4 Q9H016	Q9H016 homo sapien
10	578.5	17.6	623	5 Q9NGX7	Q9NGX7 drosophila
11	578.5	17.2	623	5 Q9VU05	Q9VU05 drosophila
12	565.5	17.2	591	11 Q8VCK5	Q8VCK5 mus musculu
13	537	16.3	593	4 Q8TBH5	Q8TBH5 homo sapien
14	526.5	16.0	586	4 Q9BQF8	Q9BQF8 homo sapien
15	526.5	16.0	744	5 Q9VENS	Q9VENS drosophila
16	524	15.9	748	4 Q8TBJ7	Q8TBJ7 homo sapien

17	523	15.9	582	4 Q96RF4	Q96RF4 homo sapien
18	517.5	15.7	564	4 Q96MV2	Q96MV2 homo sapien
19	513.5	15.6	589	4 Q96L69	Q96L69 homo sapien
20	512.5	15.6	584	4 Q96FT8	Q96FT8 homo sapien
21	509.5	15.5	586	11 Q9CZP4	Q9CZP4 mus musculu
22	504.5	15.3	707	4 Q96CT2	Q96CT2 homo sapien
23	500	15.2	568	4 Q96H35	Q96H35 homo sapien
24	499	15.2	571	11 Q9CR40	Q9CR40 mus musculu
25	498.5	15.1	568	11 Q8R2H4	Q8R2H4 rattus norv
26	498	15.1	625	5 Q9359	Q9359 drosophila
27	498	15.1	654	5 Q8SZY0	Q8SZY0 drosophila
28	497.5	15.1	531	5 Q61795	Q61795 caenorhabd
29	492	14.9	624	4 Q9BPY9	Q9BPY9 homo sapien
30	484.5	14.7	627	5 Q9VK21	Q9VK21 drosophila
31	480.5	14.6	575	5 Q9VGE5	Q9VGE5 drosophila
32	474	14.4	617	4 Q8RC02	Q8RC02 homo sapien
33	473	14.4	608	11 Q9DA07	Q9DA07 mus musculu
34	473	14.4	608	11 Q9DSV2	Q9DSV2 mus musculu
35	472.5	14.4	614	4 Q96MCO	Q96MCO homo sapien
36	470.5	14.3	610	4 Q8WZ60	Q8WZ60 homo sapien
37	467.5	14.2	589	11 Q8R2P1	Q8R2P1 mus musculu
38	466.5	14.2	589	4 Q9H0H3	Q9H0H3 homo sapien
39	460.5	14.0	538	5 Q9VGE6	Q9VGE6 drosophila
40	460	14.0	621	11 Q9D783	Q9D783 mus musculu
41	455	13.8	604	4 Q96Q17	Q96Q17 homo sapien
42	453	13.8	634	11 Q99JN2	Q99JN2 mus musculu
43	452	13.7	634	4 Q9H511	Q9H511 homo sapien
44	451	13.7	613	11 Q8RI24	Q8RI24 mus musculu
45	448	13.6	638	11 Q9DBY7	Q9DBY7 mus musculu

## ALIGNMENTS

### RESULT 1

Q9P239	PRELIMINARY;	PRT;	511 AA.
AC Q9P239;			
DT 01-OCT-2000 (TREMBLrel. 15, Created)			
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE KIAA1489 protein (Fragment).			
GN KIAA1489.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX NCBI_TaxID=9606;			
RN 11			
RP SEQUENCE FROM N.A.			
RX MEDLINE=20277482; PubMed=10819331;			
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;			
RT "Prediction of the coding sequences of unidentified human			
RT genes.XVII. The complete sequences of 100 new cDNA clones from brain			
RT which code for large proteins in vitro."			
RL DNA Res. 7:143-150(2000).			
DR EMBL; AB040922; BAA96013.1; -			
DR InterPro; IPR001798; Kelch.			
DR Pfam; PF01344; Kelch; 3.			
FT NON_TER			
FT NON_TER			
SQ SEQUENCE 511 AA; 58516 MW; 96837D756D242DBA CRC64;			

Query Match 82.7%; Score 2732; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 1, 1e-221; Indels 0; Gaps 0;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113	VEDVLQRCREYLIKINENCVRLISFADLFSCEELKQSAKRMVHKFTAVYHODAFMQL	172
DB 1	VEDVLQRCREYLIKINENCVRLISFADLFSCEELKQSAKRMVHKFTAVYHODAFMQL	60
QY 173	SHDLIDILSSDNLVKEETVREANMLLEYNTESSRQYSSVLSQIRIDALSTVQRA	232
DB 61	SHDLIDILSSDNLVKEETVREANMLLEYNTESSRQYSSVLSQIRIDALSTVQRA	120

233 WFOGLPNDKSVVVOGLYKSMKPKFKRLGMLTKEEMMIFIEASSENPCSLYSSVCYSPQA  
292  
Db WFOGLPNDKSVVVOGLYKSMKPKFKRLGMLTKEEMMIFIEASSENPCSLYSSVCYSPQA 180  
293 EKYYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHSKTSKLOTAFTVNCFTW 352  
Db EKYYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHSKTSKLOTAFTVNCFTW 240  
353 FDAQNTWPFKTMLEFVRIKPSLVCEGYIYAIGDSVGGELNRRRTVRYDTEKDMTW 412  
Db FDAQNTWPFKTMLEFVRIKPSLVCEGYIYAIGDSVGGELNRRRTVRYDTEKDMTW 300  
413 SPLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAMQTSRSPASAAAFDCKIF 472  
Db SPLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAMQTSRSPASAAAFDCKIF 360  
473 YIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI PARYSDPCRAVVISN 532  
Db YIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI PARYSDPCRAVVISN 420  
533 SLCPVARETHLNERAKYVYQYDLELRWELRHSERVLMDGRDPRCTVGLYPSCLE 592  
Db SLCPVARETHLNERAKYVYQYDLELRWELRHSERVLMDGRDPRCTVGLYPSCLE 480  
593 ESPWKPPTYLFTDGTBEFELDGEVVALPV 623  
Db ESPWKPPTYLFTDGTBEFELDGEVVALPV 511

## RESULT 2

09R1W9 PRELIMINARY; PRT; 469 AA.  
AC 09R1W9;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Hypothetical 53.6 kDa protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EYE;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC022962; AAH22962.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 469 AA; 53590 MW; 94A39E15DBCAB844 CRC64;

Query Match 75.6%; Score 2490; DB 11; Length 469;  
Best Local Similarity 99.1%; Pred. No. 4.2e-202;  
Matches 465; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

155 MEHKTAYVHQAFAQLSHDLIDILSDNLNVEKEETVREAAMLLENTESRQYLS 214  
Db MEHKTAYVHQAFAQLSHDLIDILSDNLNVEKEETVREAAMLLENTESRQYLS 60  
215 SVLSQIRIDALSVTRAWFQGLPNDKSVVVOGLYKSMKPKFKRLGMLTKEEMMIFIEA 274  
Db SVLSQIRIDALSVTRAWFQGLPNDKSVVVOGLYKSMKPKFKRLGMLTKEEMMIFIEA 120  
275 SSENPCSLYSSVCYSPQAEKVYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHS 334  
Db SSENPCSLYSSVCYSPQAEKVYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHS 180  
335 SKTSKLOTAFTVNCFTWFDQAQNTWFPKTPMLFVRIKPSLVCEGYIYAIGDSVGGEL 394  
Db SKTSKLOTAFTVNCFTWFDQAQNTWFPKTPMLFVRIKPSLVCEGYIYAIGDSVGGEL 240  
395 NRTVRYDTEKDMTWVSLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAM 454  
Db NRTVRYDTEKDMTWVSLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAM 300

455 ROTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI 514  
Db ROTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI 360  
515 IPAKYSDPCRAVVISNLSLVFMRETHLNERAKYVYQYDLELRWELRHSERVLMD 574  
Db IPAKYSDPCRAVVISNLSLVFMRETHLNERAKYVYQYDLELRWELRHSERVLMD 420  
575 LGDRFRCVGLYPSCLEESPWKPTLYLFTDGTBEFELDGEVVALPV 623  
Db LGDRFRCVGLYPSCLEESPWKPTLYLFTDGTBEFELDGEVVALPV 469

## RESULT 3

09UFM7 PRELIMINARY; PRT; 348 AA.  
AC 09UFM7;  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Hypothetical 39.5 kDa protein (Fragment).  
GN DKFZ566C134.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Bloeker H.; Boecker M.; Brandt P.; Nemes H.W.; Gassenhuber J.;  
RA Wiemann S.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL115562; CAB55994.1; -.  
DR InterPro: IPR001798; Kelch.  
DR Pfam: PF01344; Kelch; 3.  
KW Hypothetical protein.  
FT NON TER  
SQ SEQUENCE 348 AA; 39503 MW; C129A7A6CAE01A77 CRC64;

Query Match 57.3%; Score 1886; DB 4; Length 348;  
Best Local Similarity 99.4%; Pred. No. 3.5e-151;  
Matches 346; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

276 SENPCSLYSSVCYSPQAEKVYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHS 335  
Db SENPCSLYSSVCYSPQAEKVYKLCSPADLHKVGTVPNDIYIAGQVPLKTKTKNHS 60  
336 KTSKLOTAFTVNCFTWFDQAQNTWFPKTPMLFVRIKPSLVCEGYIYAIGDSVGGELN 395  
Db KTSKLOTAFTVNCFTWFDQAQNTWFPKTPMLFVRIKPSLVCEGYIYAIGDSVGGELN 120  
396 RRTVRYDTEKDMTWVSLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAMR 455  
Db RRTVRYDTEKDMTWVSLPCAMQMSAAVVDHDIYVMTLIMCYFPRSDSWEMAMR 180  
456 QTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI 515  
Db QTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDSSVVEIYDVNKNEMKMANI 240  
516 PAKYSDPCRAVVISNLSLVFMRETHLNERAKYVYQYDLELRWELRHSERVLMDL 575  
Db PAKYSDPCRAVVISNLSLVFMRETHLNERAKYVYQYDLELRWELRHSERVLMDL 300  
576 GRDFRCVGLYPSCLEESPWKPTLYLFTDGTBEFELDGEVVALPV 623  
Db GRDFRCVGLYPSCLEESPWKPTLYLFTDGTBEFELDGEVVALPV 348

## RESULT 4

09Y382 PRELIMINARY; PRT; 335 AA.  
AC 09Y382;  
ID 09Y382;  
AC 09Y382;

```

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CGI-73 protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
RT Caenorhabditis elegans by Comparative Proteomics.";
RL Genome Res. 10:703-713(2000).
DR EMBL; AF151831; AAD34068.1; -
DR InterPro: IPR000210; BTF_POZ.
DR Pfam; PF00651; BTF; 1.
DR SMART; SM00225; BTF; 1.
DR PROSITE; PS50097; BTF; 1.
SQ SEQUENCE 335 AA; 38339 MW; C1B10D6291736A44 CRC64;

Query Match 50.7%; Score 1670; DB 4; Length 335;
Best Local Similarity 98.5%; Pred. No. 6,2e-133;
Matches 326; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTDORQINTQYAVSLAEQLKLFYEQQLFTDVLIVLVEGTEPFCMKVLTATCSYFRAMF 60
DB 1 MSTDORQINTQYAVSLAEQLKLFYEQQLFTDVLIVLVEGTEPFCMKVLTATCSYFRAMF 60
QY 61 MSGLESQKQTHVLRNVDATLQIIITVYTGNLANDSTVQLYETACFLQVEDVLRQC 120
DB 61 MSGLESQKQTHVLRNVDATLQIIITVYTGNLANDSTVQLYETACFLQVEDVLRQC 120
QY 121 RRYLLKRIANECVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLLIDI 180
DB 121 RRYLLKRIANECVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLLIDI 180
QY 122 RRYLLKRIANECVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLLIDI 180
DB 122 RRYLLKRIANECVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLLIDI 180
QY 181 LSSDNLNVEKEETVEAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPN 240
DB 181 LSSDNLNVEKEETVEAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPN 240
QY 181 PSSDNLNVEKEETVEAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPN 240
DB 181 PSSDNLNVEKEETVEAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPN 240
QY 241 DKSVVVQGLYKSMKPKFKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
DB 241 DKSVVVQGLYKSMKPKFKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
QY 241 DKSVVVQGLYKSMKPKFKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
DB 241 DKSVVVQGLYKSMKPKFKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCS 300
QY 301 PRADLHKVGTVTTPNDIYIAGQVPLKNTK 331
DB 301 PRADLHKVGTVTTPNDIYIAGQVPLKNTK 331

RESULT 5
ID Q9P097 PRELIMINARY; PRT; 351 AA.
AC Q9P097;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE HSPC284 (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RA Ye H.-Y., Zhang Q.-H., Zhou J., Shen Y., Wu X.-Y., Guan Z.-Q., Wang L.,
RA Fan H.-Y., Mao Y.-F., Dai M., Huang Q.-H., Chen S.-J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161402; AAF28962.1; -
DR InterPro: IPR001798; Kelch.
DR Pfam; PF01344; Kelch; 2.

```

```

FT NON TER 1
SQ SEQUENCE 351 AA; 39256 MW; B6D8567A34CDA7DE CRC64;

Query Match 50.6%; Score 1664.5; DB 4; Length 351;
Best Local Similarity 92.1%; Pred. No. 1.9e-132;
Matches 326; Conservative 1; Mismatches 20; Indels 7; Gaps 4;

QY 194 VREAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPNPKSVVVOGLYKSM 253
DB 1 VREAAMLMLEYNTESRQYLSVLSQIRIDALSEVQORAMFQGLPNPKSVVVOGLYKSM 60
QY 254 PKFFPKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCSPPADLHKVGTVT 313
DB 61 PKFFPKRLGKTEKEMMIFIEASSENPCSLYSVCYSPQAEKYKCSPPADLHKVGTVT 119
QY 314 PNDIYIAGQVPLKNTKTHNSKTSKQTARTNCFWPAQOONTWPP--KTPMLFRI 371
DB 120 PNDIYIAGQVPLKNTKTHNSKTSKQTARTNCFWPAQOONTWPP--KTPMLFRI 175
QY 372 KPSLVCCGTYIATGDSVSGELNRRRTVERDYDEKDEWTWVSPJPCAMQMSAAVVDHI 431
DB 176 KAIFCCGCGTYIATGDSVSGELNRRRTVERDYDEKDEWTWVSPJPCAMQMSAAVVDHI 235
QY 432 YVMTLNTMYCYFPNSDSVTEMAMROTSSRFASAAAFGDKIYIGLHATNSGIRLPST 491
DB 236 YVMTLNTMYCYFPNSDSVTEMAMROTSSRFASAAAFGDKIYIGLHATNSGIRLPST 295
QY 492 VDGSSVTEIYDVNKNEMKMANI PAKRYSDPCYAVVISLGFVMEETHLNE 545
DB 296 VDGSSVTEIYDVNKNEMKMANI PAKRYSDPCYAVVISLGFVMEETHLNE 349

RESULT 6
ID Q96015 PRELIMINARY; PRT; 526 AA.
AC Q96015;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE KIAA1842 protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058745; BAB47471.1; -
DR InterPro: IPR000210; BTF_POZ.
DR InterPro: IPR001798; Kelch.
DR Pfam; PF00651; BTF; 1.
DR Pfam; PF01344; Kelch; 3.
DR PROSITE; PS50097; BTF; 1.
DR NON TER 1
SQ SEQUENCE 526 AA; 60813 MW; 8B0A24BD234A5970 CRC64;

Query Match 23.6%; Score 778.5; DB 4; Length 526;
Best Local Similarity 32.0%; Pred. No. 3.3e-57;
Matches 176; Conservative 102; Mismatches 233; Indels 39; Gaps 9;

QY 57 RAMFSGSESKQTHVLRNVDATLQIIITVYTGNLANDSTVBCVETACFLQVEDV 116
DB 1 RAMFSGSESKQTHVLRNVDATLQIIITVYTGNLANDSTVBCVETACFLQVEDV 60
QY 117 LQRCREYLIKINNEVCVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLD 176
DB 117 LQRCREYLIKINNEVCVRLSPADLPSCBELKQSAKRWKHECTAVYHODAMQSHLD 176

```

```

Db      61  QDQCAKYMISHLDPONSIGVFIADHYHQBLGDRSKKEYIRKKFLCVTEQOEFLQLTQKQ 120
Qy      177  LIDILSDNLNVEKEETVREAAAMLLEYNTSSROYLSVLSQ-IRIDALSEVTOBAMPQ 235
Db      121  LILISLSDNLNVEKEETVREAAAMLLEYNTSSROYLSVLSQ-IRIDALSEVTOBAMPQ 235
Qy      236  GLPNDKSVVVOGLYKSMKPKFK---PRLQMTKEEMIFLEASSENPCSLYSVCYSPQA 292
Db      177  KIPFOFAQIAKSCVEKGPNTNGCTQRLQMTASEMIICFDAHKHSGKKQVPCLDIYV 236
Qy      293  EKYYKCSPPADLHKVTVTPNDIYIAGQVPLKTKTNHKSLSKLOTAFTVNCFTV 352
Db      237  GRVFKLCKPPNDLREGVILVSPNDIYIAGQVPLKTKTNHKSLSKLOTAFTVNCFTV 287
Qy      353  FDAQNTWPFKPTMFLFKIKPSLVCEGYIYAGDSVSGE--LNRRTVERVYDEKDEMT 410
Db      288  YDISTRNMLSKPRLRIRKICCKLVYCGKMYALGKRYEGEDGNSLSKVCYCSRENCMT 347
Qy      411  MVSPLPCAMQMSAAVVVHDCIYVMTLNLAMYCFPRSDSVWEMAMROTSSPASAAGDK 470
Db      348  TVCAMPAVAMEFNHNAVEKEKIVYLOGEFFLEFEPQDYMGLTPMTVPRIQGLAAVYKDS 407
Qy      471  IPIFGHLHTNGIRLPSGTVDSSVTVIIVYVKNKEMMAANIPAKRISDPCVAVYI 530
Db      408  IYIYIAG---TCGNHQRN-----FTVBAVDIELNKWTRKKDPDQDSINPYLKLVL 455
Qy      531  SNLSLCVFMRETHL-----NERAKVTVYQDLELDRLWSLRQHSERVLMDLGRDPCYV 583
Db      456  QNTLHLFVRAITQVTEVHEVTRTSKSLYQYDIDIAQMKMYETPPDR-LMDLGRHECAV 514
Qy      584  GKLYPSCLEB 593
Db      515  AKLYPQCLQK 524

RESULT 7
Q8MWZ9  ID  Q8MWZ9  PRELIMINARY;  PRT;  684  AA.
AC      08MWZ9;
DT      01-MAR-2002 (TEMBLrel. 20, Created)
DT      01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DS      Hypothetical 77.2 kDa protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=TESTIS;
RA      Strausberg R.;
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL: BC022033; AAB22033.1; -
DR      InterPro: IPR000210; BTB_POZ.
DR      InterPro: IPR001798; Kelch.
DR      Pfam: PF00651; BTB; 1.
DR      Pfam: PF01344; Kelch; 3.
DR      SMART: SM00225; BTB; 1.
DR      PROSITE: PSS0097; BTB; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 684 AA; 77163 MW; B2DBB6969A1AESE9 CRC64;

Query Match 19.4%; Score 638.5; DB 4; Length 684;
Best local similarity 26.1%; Pred. No. 3.4e-45;
Matches 177; Conservative 125; Mismatches 248; Indels 127; Gaps 21;

Qy      3  TODEROINTERAVSLLEQLLFEYQQLFTDILIV-----EGTEFPCHKVAVATCSS 54
Db      35  TGBEELKDTAHSALALQKSFYDARLLCVTIEVTPSGSGPTGLFLSCNRVLAACP 94
Qy      55  YFPAAMSGLSSEKQTHVHLRANDAATLQIITVATGNLANDSVVEQLYFACLOVE 114
Db      95  YFASMTGTGMYESQOASVTMHDVDASSFEVLVDYCYTGRVSLSEANVORLYAASDMLQLE 154

```

```

Qy      115  DVLQRCREYLIIKINAENCVRLLSPADLFSCEELKQSAKRWHEKFTAV-----YHDAF 169
Db      155  YVREACASLARLRLDTLCTALIKRADAFDHHKLRQSQASYAHNFKOLSRNGSIREETL 214
Qy      170  KOLSHDLIDILSSDNLNVEKEETVREAAAMLLEYNTSSROYLSVLSQIRIDALSEVT 229
Db      215  ADLTLAQLLAVALRLSLDIESERTVCHVAQVLEAAERGPSAIEFKVCVMMHFTED 274
Qy      230  QRAMFOGL--PPNDSSV---VVOG-----LYKSM----- 253
Db      275  Q-DYLEGTLTKPIVKYCLDVLLEGALQMRYGDLTKSLVPVPNSSSSSSSSSLVSAABN 333
Qy      254  -PKFKPRLQMTKEEMIFLEASSENPCSLYSVCYSPQAQKYYLCSPPADLHKVGY- 311
Db      334  PQQ-----RLGMAKEMVIFP-GHRPD-----FLCYDYSQDLYTMPSPPLSFAHTKYV 383
Qy      312  ----VTPNDIYIAGQVPLKTKTNHKSLSKLOTAFTVNCFPVPAQNTWPKPKPM 366
Db      384  SSAVCVSPDHDIYLA--QPRKD-----LWVYKPAQNSMOQLADR 421
Qy      367  LFVRIKPSLVCEGYIYAG--DSVGGELNRTVERVYDEKDEMTMVSPLPCAMQMSAAV 425
Db      422  LLCREGMDVAYINGIYILGGRDPIYG-VKLEVECYVQRNQMALVAPVPSFYSFELI 480
Qy      426  VVHDCIYVMTLNLAMYCFPRSDSVWEMAMROTSSPASAAGDKIPIYIGLHITNSGI 485
Db      481  VQNTLVAVNSKRMCLCYPSHNMWMLNCASLRS-DFOGACVFNDEIYICDIPV----- 533
Qy      486  RLPSGTVDSSVTVIIVYVKNKEMMAANIPAKRISDPCVAVYISNLSLCVFMRETHLNE 545
Db      534  -----MKVYNPARGEMWRISNIPDSERTN-YQIVNHDQKLILLITSTTPQMK 579
Qy      546  RAKVTVYQDLELDRLWSLRQHSERVLMDLGRDPCYTGKLYPSCLESMPKPYTLST 605
Db      580  KNRTVYETREDQMINIGTWLGLQPDG--FICLCARVPSCLB-----PQGSFIT 631
Qy      606  -----DGTBEFELDQ 615
Db      632  EEDDARSSSTEMWLDG 648

RESULT 8
Q96MP6  ID  Q96MP6  PRELIMINARY;  PRT;  674  AA.
AC      Q96MP6;
DT      01-DEC-2001 (TEMBLrel. 19, Created)
DT      01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE      CDNA FLJ32071 f1e, clone OCBF1000122, weakly similar to RING CANAL
DE      protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=BRNIN;
RA      Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA      Arita M., Mueashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA      Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA      Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA      Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA      Magatsuna M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
RA      Suzuki Y., Sugano S., Negahari K., Masuno Y., Nagai K., Isogai T.,
RT      "NEO human cDNA sequencing project."
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AK056633; BAB71238.1; -
DR      InterPro: IPR000210; BTB_POZ.
DR      InterPro: IPR001798; Kelch.
DR      Pfam: PF00651; BTB; 1.
DR      Pfam: PF01344; Kelch; 1.
DR      PROSITE: PSS0097; BTB; 1.

```

SQ SEQUENCE 674 AA; 76152 MW; 98275596E094AFB CRC64;  
 Query Match 19.2%; Score 631.5; DB 4; Length 674;  
 Best Local Similarity 26.1%; Pred. No. 1.3e-44;  
 Matches 177; Conservative 128; Mismatches 244; Indels 129; Gaps 22;  
 QY 3 TQDERQINTEYAVSLLEQLKLFYEQQLFTDVLIV-----EGTEPCHKKVLTATCSS 54  
 DB 35 TQPEELKDTAHSALLAQLKSFYDARQLCDVTIEVTPGSGPGTGRLFPCNNVLAACP 94  
 QY 55 YFRAMFMSGLSESKQTHVLRNDATLQIITVATGNLANNDSVEQLYETACFLQVE 114  
 DB 95 YFKSMFTGMYESQASVTMHVDVASFEVLVDYCTGVSLSEANVERLYAASDMLQLE 154  
 QY 115 DYLRQCREYLKIKINAEVCRLSPADLFSCBELKQSAKRWENHKTAVYH-----QDAF 169  
 DB 155 YREACASFLARLDLTNCTAILKFPDAGHRLKRSQASQSYIAHNFQSLRWGSIREEFL 214  
 QY 170 MQLSHDLIDLISDNLVKEETVEBAAMLLEVENTESROYLSSYLQIRIDALSEVT 229  
 DB 215 ADLTIAQLLAVLRDLSDVSEEGTCHVAVQWLEAPKRGPSAAEVFCVAMHPTFED 274  
 QY 230 QRAMFGGL--PNDKSV---VVOG-----LYKSM----- 253  
 DB 275 Q-DYLEGLLTKPIYKCYCLDVIIEGALQMRGYDLVPSVPNPNSSSSSSSLVSAEN 333  
 QY 254 -PKFPRKLGMTKEEMKFIIEASSENPCSLYSVCYSPQAEKYKLCSPPADLHKVTV 312  
 DB 334 PFGQ---RLGMCACKEMVIFP-GHPRDP-----FLCDDPYSGDLYKVPSPFLTCLAHRTVT 383  
 QY 313 T-----PNDNIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFFWFDAAQONTWPKTPM 366  
 DB 384 TLAVCISPDHDIYLA--QPRD-----LMVYKPAQNSMOQLADR 421  
 QY 367 LVVRIPSLVCCCEGYIYAGG-DSVGGELNRTVERVYDEKDEMTVSPPLPCAMQMSAAV 425  
 DB 422 LRCRGMADVAYLNGYIYILGGRDPTG-VKLKEVECYVQRNOMALVAPVPSFISFLDM 480  
 QY 426 VVHDCIYVMTLNLMYCFFPRSDSVWEMAMRQTSRSPASAAFGDKIFYGGLHAIATNGI 485  
 DB 481 VVRDIYALNSKRMFCYDPSHNMMLKCVSLKRN-DFOEACVFNEEYICIDIPV----- 533  
 QY 486 RLPSGTVDSSVTVEIYVNNKNEWMAANIPAKRYSDPCRAVVISNLSLCEVMEETHANE 545  
 DB 534 -----MKYVNPVRAEWQNNIPLVSETHN--RIIKHGQKLILTSRTPQMK 579  
 QY 546 RAKYVTYQYDLELDKRSLSRQHSERV-LMDLGHDFCTGKGLYPSGLSESPMKPPTLYFS 604  
 DB 580 KKRVTYVEYDTRDDQWI---NIGTTLGLQFDSNFCLSARVYPSCLE-----PQGSFL 630  
 QY 605 TDGTE-----EFELDG 615  
 DB 631 TEEBEPSESTEWDLGG 648  
 RESULT 9  
 Q9H016 PRELIMINARY; PRT; 684 AA.  
 AC Q9H016 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 77.2 kDa protein.  
 GN DKFZP434E2318.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RX MEDLINE=21154917; PubMed=11330166;  
 RA Wiemann S., Weill B., Wellenreuther R., Gaassenhuber J., Glasel S.,

RA Ansgorge W., Boecker M., Bloecker H., Bauersache S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Struck N.,  
 RA Mewes H.W., Oettermeyer B., Obermaier B., Tampe J., Heubner D.,  
 RA Wambutt R., Korn B., Klein M., Poustka A.,  
 RA "Towards a Catalog of Human Genes and Proteins: Sequencing and  
 RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."  
 RL Genome Res. 11:422-435(2001).  
 DR EMBL, AL136782; CAB66716.1; -  
 DR InterPro; IPR000210; BTF POZ.  
 DR InterPro; IPR001798; Kelch.  
 DR Pfam; PF00651; BTF; 1.  
 DR Pfam; PF01344; Kelch; 3.  
 DR SMART; SM00225; BTF; 1.  
 DR PROSITE; PS50097; BTF; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 684 AA; 77153 MW; 9DD9273515A77555 CRC64;  
 Query Match 19.2%; Score 630.5; DB 4; Length 684;  
 Best Local Similarity 26.0%; Pred. No. 1.6e-44;  
 Matches 176; Conservative 125; Mismatches 249; Indels 127; Gaps 21;  
 QY 3 TQDERQINTEYAVSLLEQLKLFYEQQLFTDVLIV-----EGTEPCHKKVLTATCSS 54  
 DB 35 TQPEELKDTAHSALLAQLKSFYDARQLCDVTIEVTPGSGPGTGRLFPCNNVLAACP 94  
 QY 55 YFRAMFMSGLSESKQTHVLRNDATLQIITVATGNLANNDSVEQLYETACFLQVE 114  
 DB 95 YFKSMFTGMYESQASVTMHVDVASFEVLVDYCTGVSLSEANVERLYAASDMLQLE 154  
 QY 115 DYLRQCREYLKIKINAEVCRLSPADLFSCBELKQSAKRWENHKTAVYH-----YHDAF 169  
 DB 155 YREACASFLARLDLTNCTAILKFPDAGHRLKRSQASQSYIAHNFQSLRWGSIREEFL 214  
 QY 170 MQLSHDLIDLISDNLVKEETVEBAAMLLEVENTESROYLSSYLQIRIDALSEVT 229  
 DB 215 ADLTIAQLLAVLRDLSDVSEEGTCHVAVQWLEAPKRGPSAAEVFCVAMHPTFED 274  
 QY 230 QRAMFGGL--PNDKSV---VVOG-----LYKSM----- 253  
 DB 275 Q-DYLEGLLTKPIYKCYCLDVIIEGALQMRGYDLVPSVPNPNSSSSSSSLVSAEN 333  
 QY 254 -PKFPRKLGMTKEEMKFIIEASSENPCSLYSVCYSPQAEKYKLCSPPADLHKVTV 311  
 DB 334 PFGQ---RLGMCACKEMVIFP-GHPRDP-----FLCYDSYSGDLYTMPSPFLTSAHTKTVT 383  
 QY 312 -----VTDNIDYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFFWFDAAQONTWPKTPM 366  
 DB 384 SSAVCSPDHDYLA--QPRD-----LMVYKPAQNSMOQLADR 421  
 QY 367 LVVRIPSLVCCCEGYIYAGG-DSVGGELNRTVERVYDEKDEMTVSPPLPCAMQMSAAV 425  
 DB 422 LRCRGMADVAYLNGYIYILGGRDPTG-VKLKEVECYVQRNOMALVAPVPSFISFLI 480  
 QY 426 VVHDCIYVMTLNLMYCFFPRSDSVWEMAMRQTSRSPASAAFGDKIFYGGLHAIATNGI 485  
 DB 481 VVQNYLVAVNSKRMFCYDPSHNMMLKCVSLKRN-DFOEACVFNEEYICIDIPV----- 533  
 QY 486 RLPSGTVDSSVTVEIYVNNKNEWMAANIPAKRYSDPCRAVVISNLSLCEVMEETHANE 545  
 DB 534 -----MKYVNPARGEWRRISNIPLSETHN-YQIVNHDQKLILTSRTPQMK 579  
 QY 546 RAKYVTYQYDLELDKRSLSRQHSERV-LMDLGHDFCTGKGLYPSGLSESPMKPPTLYFS 605  
 DB 580 KKRVTYVEYDTRDDQWI---NIGTTLGLQFDSNFCLSARVYPSCLE-----PQGSFL 631  
 QY 606 -----DTEBEFLDG 615  
 DB 632 EEDDARSSSTEWDLGG 648  
 RESULT 10  
 Q9NGX7 PRELIMINARY; PRT; 623 AA.  
 ID Q9NGX7

AC G9NGX7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Diablio.  
 GN DBO OR CG6224.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxId=7227;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Stuart B.D., Masserman S.A.;  
 RT "Identification of Diablio, a new Drosophila melanogaster Kelch family  
 RT protein";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF237711; AAF43447.1; -.  
 DR FlyBase; FBgn0040230; dbo.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR01798; Kelch.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF01344; Kelch; 6.  
 DR PRINTS; PRO0501; KELCHREPEAT.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PSS0097; BTB; 1.  
 SQ SEQUENCE 623 AA; 68951 MW; E65817EFDEF4E7B6 CRC64;

Query Match 17.6%; Score 578.5; DB 5; Length 623;  
 Best Local Similarity 27.4%; Pred. No. 3.5e-40;  
 Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;

18 LEQKLFEYEQQLFTDIVLIVGTEFPCHKVNLATCSSYFRAMFMGSLSESKQTHVHLNV 77  
 59 LTELNMRLRRELCDVLVNGRKIFAHRYLSACSSYFCAMFTGELSEBROQEVITRDI 118  
 78 DAATLQIITVAYTGNLAMDSTVEQLYETACPLQVEDVLQRCREYLKIKINANCVRLL 137  
 119 DENAMELLIDFCYTAHIVEESNVQTLPAACLLQIVEIODICEFLKQLDPTNCLIGIR 178  
 138 SPADLFSCEELKQSAKMEVHKFTAYVHODAFQSLDILLIDLSNINLVEKEEYVREA 197  
 179 AFDYTHSCRELHIDAFTHNQVEWSESEFLLPVGLVDIICSEDLNVRSEDEVNA 238  
 198 AMLELVNTERSSQYSSVLSQIRIDALS-----EVTQRAWFOG 236  
 239 VMGLKKNVNAERQHLAQVQHRLPLSLKPLVGVTSLLVRSDEVCDDLVDKAKNYL 298  
 237 LRPNDKSVVQGLYKSNPKFKKPLKMTKEEMMIFLEASSENPCSLYSVCSYSPQAEKYV 296  
 299 LLEQERF-LMQG--PRTRPR--KP---TRGEVLFAVGCGSGDALASVERFDPQTNND-W 349  
 297 KLSPPADLHKVGVTVTPND--IYINGGVPLKTKTKNSKTKLQTAFTVNCFFWFA 355  
 350 KMAVAFNSK-RRCGGAVNLMDLLYAVGG-----HGGSYLNSIER-----YDP 391  
 356 QQNTW-FPKPTMFLVRIKPSLVCCEGYIYAGDSVGGELNRRTRVERDYDEKEMTWSP 414  
 392 QTNQMSCDVAPFTTSCRSVGVAVLDGLYAVGGDQVQCLNH--VERIDYDEKEMKSNVAP 449  
 415 LPCAMQMSAAVVVHDCIYV-----TLNLMVCPYPRSDSVEMANRQTSRSPASAAAF 467  
 450 MTRRLRGVAVAVLGGFLYAGSDGCPNLTVERYPDRHKVAVSPMSTRHGLCAVAF 509  
 468 GDKIIFYG-----GLHATNSGIRLSSGTVDS 495  
 510 NNTIYAVGGDDCMELSSAERYNPLTNWSPRYAMTSRRSGVLAVVNGQLYAVGGFDGS 569  
 496 SV--TVEIYDVNKNEMKMANIPAKR 519  
 570 AYLKTIIEVDYDPETNQRKLCGMNYRR 595

RESULT 11  
 Q9YU5  
 ID Q9YU5 PRELIMINARY; PRT; 623 AA.  
 AC Q9YU5;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CG6224 protein.  
 GN DBO OR CG6224.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxId=7227;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Suton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champs M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benze P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockett P., Brottier P.,  
 RA Burts D.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dantke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merilov G., Milshina N.V., Modarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relvert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Scapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.W., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195 (2000).  
 DR EMBL; AE003529; AAF49578.1; -.  
 DR FlyBase; FBgn0040230; dbo.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001798; Kelch.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF01344; Kelch; 6.  
 DR PRINTS; PRO0501; KELCHREPEAT.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PSS0097; BTB; 1.  
 SQ SEQUENCE 623 AA; DCD75B9F66BDC6B9 CRC64;

Query Match 17.6%; Score 578.5; DB 5; Length 623;  
 Best Local Similarity 27.4%; Pred. No. 3.5e-40;  
 Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;

18 LEQKLFEYEQQLFTDIVLIVGTEFPCHKVNLATCSSYFRAMFMGSLSESKQTHVHLNV 77

Db 59 LTELANKRHRRLCDVAVLVGKRIFAHVIILSACSSYFCAMFTGELLESROTETIRDI 118  
Qy 78 DAATLQIIITTAATGNLANNDSTVEQLYETACFLQVEDVYQRCREYLIRKINAENCVRLL 137  
Db 119 DENAMELLIDFCYTAHIIVEESNVQTLPLPAACLLQVEI IDICCEFLKQDLPTNLCGR 178  
Qy 138 SPADLFSCEELKQSAKRWHEKFTAAVYHODAFMQLSHDLIDILSDNLNVEKEEYVRA 197  
Db 179 ARAPTHSCRELLRIADKFTQHNQFWESEEFLLLVGQLVDIICSDENLVSEEOVFNA 238  
Qy 198 AMULEYNTESQYLSVLSQIRIDALS-----EYVQRAFMQ 236  
Db 239 VMSMLKYNABERQHLAQLVQHLPLSPKFLVGTGSDLVRSDEACRDVDEAKNYL 298  
Qy 237 LPENDKSVVVOGLYKMPKFKRLQMTKEEMMIFLEASSENPCSLYSSVCSPOAEKY 296  
Db 299 LIPQER- LMQG- PRIRPR- KP-----TRGEVLFAVGCGSGDALASVBRPQIND- W 349  
Qy 297 KLCSPPADLHKVGTVTTPDND- IYIAGGVPLKNTNTSKTSKLTAPFTVNCFYWPA 355  
Db 350 KVAAPMSK- RRCGVAVLNDLILYAVG-----HDGQSYLNSIER-----YDP 391  
Qy 356 QONTW- FPKTPMLFVRIKPSLVCEGVIYAGSDVSGELNRTVERDYDEKDEMTVP 414  
Db 392 QTNQMSCDVAFPTTSCRTSVGAVLDGFLYAVGSDGVQCLNH- VERYPKEMKSKVAP 449  
Qy 415 LPCAMQSAVAVVHDCIYVW-----TLNMYCYPREDSVWEMARQTSRFSASAAP 467  
Db 450 MTRTRGVAVVAVLGGFLVLAIGSDGCPNTVERYPHNKVAVSPMSTRHGLCANF 509  
Qy 468 GDKIPIFYG-----GLHIATNSGIRLPSGTVDG 495  
Db 510 NNYIYAVGGRDQCMELSSAERYNPLTNTWSPIVAMTSRRSGVLAVNGQLYAVGFGDS 569  
Qy 496 SV--TVEIYDVNKNEMKMANIPAKR 519  
Db 570 AYLKTIETVDPETNOMRLCGCMYRR 595

## RESULT 12

Q8VBKS PRELIMINARY; PRT; 591 AA.  
ID Q8VBKS  
AC Q8VBKS  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 65.9 kDa protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC019571; AAH19571.1; -  
DR InterPro; IPR000210; BTR POZ.  
DR InterPro; IPR001798; Kelch.  
DR InterPro; IPR001005; MYD\_DNA\_binding.  
DR Pfam; PF00651; BTR 1.  
DR Pfam; PF01344; Kelch; 6.  
DR PRINTS; PRO0501; KLCHESEPEAT.  
DR SMART; SM00225; BTR 1.  
DR PROSITE; PS00097; BTR 1.  
DR PROSITE; PS00334; MYB\_2; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 591 AA; 65939 MW; D4A737PAD7P1A2EB CRC64;

Query Match 17.2%; Score 565.5; DB 11; Length 591;  
Best Local Similarity 26.3%; Pred. No. 4, 1e-39;  
Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

Qy 9 INTEVAVSLBOLKLFYEQQLFTDIVLYVEGTEPFCMKVLTATGCSYFRAMSGLSBSK 68  
Db 28 IDKPRQTELEVINILRKRLRELCDVAVLVGAKKIFAHVIILSACSSYFRAMFTGELABSR 87  
Qy 69 QTHVLRNVDAATLQIIITTAATGNLANNDSTVEQLYETACFLQVEDVYQRCREYLIRK 128  
Db 88 QTEVIRIDDERAMELLIDFATTSQIYEEGNVQTLPAACLLQVEI IDICCEFLKQDLPTNLCGR 147  
Qy 129 NANCVRLLSPADLFSCEELKQSAKRWHEKFTAAVYHODAFMQLSHDLIDILSDNLN 188  
Db 148 DSNCLGIRAFADTHSCRELLRIADKFTQHNQFWESEEFLLLVGQLVDIICSDENLV 207  
Qy 189 EKEETVREAMULEYNTESQYLSVLSQIRIDALS-----E 227  
Db 208 RSEEOVFNAVMVAVKYSIQERRPOLPYQLVHRLPLSPKFLVGTGSDLVRSDEACRD 267  
Qy 228 VQGRAMFQGLPENDKSVVVOGLYKMPKFKRLQMTKEEMMIFLEASSENPCSLYSSV 284  
Db 268 LVDEAKNYLLIPQER- LMQG- PR- TRPKKIRCEVLFAVGW---CSGDALSS 316  
Qy 285 SVCSYPOAEKYKLCSPPADLHKVGTVTTPDNDIYIAG-----QVPLKNTKTNH-- 334  
Db 317 VERYPQNTNEMWVMSKRGCGVSVL- DDLIYAVGHDSSYLNSTVERDPTNOMS 375  
Qy 335 ---SKTSKLQTA-----FRTVNCFYWFAQONTWPEKTPMLFVRIKP 373  
Db 376 SDVAFPTTSCRTSVGAVLGGFLYAVGSDGVQCLNIVERYPKEMKSTRLCV 435  
Qy 374 SLVCEGYIYAGSDVSGELNRTVERDYDEKDEMTVPSPCKAMQSAVAVVHDCIY 433  
Db 436 AVAVLGGFLYAVGSDGTSPLN- TVERNPEENMHTIAPGTRRKHGLCAVYODMYA 493  
Qy 434 M-----TLNMYCYPREDSVWEMARQTSRFSASAAPFQDKIFIGLHIATNSGIR 486  
Db 494 VGGRDITELSSAERYNPLTNTWSPIVAMTSRRSGVLAVNGQLYAVGFG----- 544  
Qy 487 LPSTGVDGSSV--TVEIYDVNKNEMKMANIPAKR 519  
Db 545 -----DGTTLKTIETVDPETNOMRLCGCMYRR 573

## RESULT 13

Q8TBHS PRELIMINARY; PRT; 593 AA.  
ID Q8TBHS  
AC Q8TBHS  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Kelch-like 2, Mayven (Drosophila).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC022503; AAH22503.1; -  
SQ SEQUENCE 593 AA; 66075 MW; 3000E53B69E57F0D CRC64;

Query Match 16.3%; Score 537; DB 4; Length 593;  
Best Local Similarity 25.1%; Pred. No. 1, 1e-36;  
Matches 136; Conservative 101; Mismatches 204; Indels 100; Gaps 13;

Qy 5 DERQINTB-----YAVSLBOLKLFYEQQLFTDIVLYVEGTEPFCMKVLTAT 51  
Db 20 DSKDNTTEKHGCVTVNPMHMKKAFKVMNELR---SQNLICDVTIYABDMELISAHVVLAA 76  
Qy 52 CSYFRAMMSGLSRSGKQTHVLRNVDAATLQIIITTAATGNLANNDSTVEQLYETACFL 111  
Db 77 CSPYHAFMTGMSBSRAKRVRIKEVDGWTLMILIDIVYTAIQTVERVOVLLPAGLL 136

QY 112 QVEDVLORCREYIILKINAEVCVRLISFADLFSCBELKQSAKRMVEHKFTAVYHDAFMO 171  
 Db 137 QLODVKTCCEFLSESLQHPVNCGLGIRAFADMHACTDLINKANTYAEQHFADVLSEEFIN 196  
 QY 172 LSHDLLIDILSSINLWKEETREAMLMLENTESRSQYLSVLSQIRIDALSE--VT 229  
 Db 197 LGIEOVCSLISDGLTISSEKVEAVIAVNHDKDROSEFMRLMEHVLPLPREVLY 256  
 QY 230 QRAMFOGLPND--KSVVQGL-YKSMPEFKPRLQMTKEBMIFLEASSENPCSLYS 285  
 Db 257 QRVEEALVKNSSACDYILEAKKHYLP-----TEQRILM----- 292  
 QY 286 VCSFPOAEKVKYKLCSPPADLHKVGTVPNDIYIAGGVPLKNTYNSKTSKLOTAER 345  
 Db 293 -----KSVTRLRTPMNLPLKLMVVY-----GGQAP-----KAIR 321  
 QY 346 TVNCFWFDQONTWPKPTMFLVRILKPSLVCEGYIYAGDSVGGELNRTVERDYDE 405  
 Db 322 SYEC---YDFKERHMOVAELPSRCRCAGVYAGLVFAVG--FNGSLKVRTVDSDYDPV 376  
 QY 406 KDEMTWVSLPCAMOWSAVVVHDCIYVM-----TLNLMCYFPRSDSWEMAROTS 458  
 Db 377 KDQMTSVANMRDRSTLGAVALNGLIYAVGFDGSGTGLSVEAYNIKSNMFHVAEMNTR 436  
 QY 459 RSPASAAAFGDKIFYIGGLHIATNSGIRLPSGTVDSSVTVEIYDVNKEKMAANI 518  
 Db 437 RSSVGVGVGGLIYAVGVYDVASRQCLIS-----TVECYNATYEMTYIAEMSTR 485  
 QY 519 R 519  
 Db 486 R 486

## RESULT 14

Q9BOF8 PRELIMINARY; PRT; 586 AA.  
 AC Q9BOF8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 66.0 kDa protein.  
 GN DKFZ564C16.6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=21154917; PubMed=11230166;  
 RA Mlemin S., Weil B., Wellenreuther R., Gaessenhuber J., Glaesl S.,  
 RA Ansoege W., Boescher M., Bloeker H., Bauersachs S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
 RA Mewes H.W., Ottenwelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wandt R., Korn B., Klein M., Poulska A.;  
 RA "Towards a Catalog of Human Genes and Proteins: Sequencing and  
 RT Analysis of 500 Novel Complete Protein Coding Human cDNAs";  
 RL Genome Res. 11:422-435 (2001).  
 DR EMBL: AL136597; CAB6532.1; -.  
 DR HSSP: Q05516; ICS3.  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR01798; Kelch.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF01344; Kelch; 5.  
 DR SMART: SM00325; BTB; 1.  
 DR PROSITE: PS50057; BTB; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 586 AA; 65991 MW; FA4ED8F1567AC7A8 CRC64;

Query Match 16.0%; Score 526.5; DB 4; Length 586;  
 Best Local Similarity 25.7%; Pred. No. 8.1e-36;  
 Matches 156; Conservative 100; Mismatches 229; Indels 123; Gaps 20;

QY 11 TEVAVSLLEQLTL-----FYEQQLFTDILVILEGTEPPCHRMVLATCSSYFRAME 60  
 Db 14 TEKKLAAREAKLGLFNGVMNMNRKQKTLCDVILMVOGRKIPARHVLLAAASHFEINMF 73  
 QY 61 MGSLSSKQTHHLNVDATLQIIITAYTGNLAMDSTVBOLEYTACFCGLVEVQLQC 120  
 Db 74 TTNMLESSEFEVLDAEDDILIEQLVEFYATRIIVSNNSVSLDPAANOYIEPVKKVC 133  
 QY 121 REYLTKKINAEVCVRLISFADLFSCBELKQSAKRMVEHKFTAVYHDAFMOQLSHDLLIDI 180  
 Db 134 VDFLEKQVDASNCGLSVLAECDDPELKATDDPILHQHFEVYTKDTDELQDVKRYTHL 193  
 QY 181 LSSDNLWKEETREAMLMLEINTESRSQYLSVLSQIRIDALSE----- 227  
 Db 194 LMQDILTVAEDQVYDAVRMLKYDEPNRPVMDILAKVRPLISKFLSKTVOAEPLI 253  
 QY 228 -----VTQRAPFOGLPNDKSVVQGLYKSMPEFKPRLQMTKEBMIFLEASSEN 278  
 Db 254 QDNPECLKRVISGMRYHLISPEDRELVDT-----RPRRKGDYRIALF-GGSQPO 304  
 QY 279 PCSLYSVCSYPOAEKVKYKLCSPPADLHKVGTVPNDIYIAGGO--VPLKNTYNSK 336  
 Db 305 SCRYENPKDYS-----WTDIRCPFEKRDAACVMDNVYILGSSQLFPIR----- 351  
 QY 337 TSKLOTAFTVNCFYWFDQONTWF---PKTPMLFVRILKPSLVCC--EGYIYALIGDSV 390  
 Db 352 -----MDC---YVVVKDSWYSKLGPPTR-----RDLAACAAGKITYSGSSEV 392  
 QY 391 GGLNRRYVERDYDTEKDEK-TWVSLP--CAMOWSAVVVHDCIYV-----MTL 436  
 Db 393 GNSA-LYLFECYDITRESWHTKPSMLTORCS--HGWENALIIYCGSLANNVSGRYL 448  
 QY 437 NLMYCFPRSDSWEMAROTSRSFASAAAFGDKIFYIGGLHIATNSGIRLPSGTVDGSS 496  
 Db 449 NSCEVYDPAETETWELCPMIEARKHGLVYFVQDKIFAVCG-----QNGL-----GGL 495  
 QY 497 VTVEIYDVNKEKMAANIPAKRYSDPCVRAVVISNLCVEN--RETHLNERAKYVYQ 553  
 Db 496 DNVVEYDIDIKLNMKVVSPMPKGVTKCAVGSIVYVLAGFGVGLGHILE----- 547  
 QY 554 YDLLELDRW 561  
 Db 548 YNTETDKW 555

## RESULT 15

Q9VENS PRELIMINARY; PRT; 744 AA.  
 AC Q9VENS;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE CG3962 protein.  
 GN CG3962.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bendale U., Bayraktiroglu U., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,



RA Burtis K.C., Bugam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.E., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart K.M., Glaeser K.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hoselin D., Houshun F., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Maasman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003716; AAF55386.1; -.  
DR Flybase; FBgn038475; CG3962.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR001798; Kelch.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF01344; Kelch; 6.  
DR PRINTS; PR00501; KELCHREPEAT.  
DR SMART; SM00225; BTB; 1.  
DR PROSITE; PSS0097; BTB; 1.  
SQ SEQUENCE 744 AA; 83604 MW; 6F7A4EB69DF482E9 CRC64;

Query Match 16.0%; Score 526.5; DB 5; Length 744;  
Best Local Similarity 28.0%; Pred. No. 1.2e-35;  
Matches 157; Conservative 77; Mismatches 207; Indels 119; Gaps 18;

QY 11 TEVAVSLLEQLKLYEQQLFTDVLIVESTEPCHMVLATCSSYPRAMFMSGLSESQOT 70  
DB 38 SNVAKEALQGMWMMRSHGLTDLVLEVKELFPAHKVVLASAAPYKAMFTGSLKSEMS 97  
QY 71 HVHLRNVDAATLQIITTAAYTGNLANSTVEQLYETACELQVEDVLRCREYLKIKINA 130  
DB 98 RVQLQGVCTPAMSRILYFMVYTGQIRVTEVTCQLLPAAATMFOVPNVIDACCAFLERQIDP 157  
QY 131 ENCVRLLSFADLFSCSELKQSAKRWVHEKFTAVYHODAFMOLSHDLLIDLSDNLVK 190  
DB 158 TNAIGIAHFAEOHCVELQKRAVFLERNFTQVCOEERFLQSAVOLIALIRDELDNVOE 217  
QY 191 EETVREAAMLDEYNTESRSQYLSVLSQIRIDALSEVTCRAMFOGLPPNDKSVVVOGLY 250  
DB 218 EREVVNAVLKMYVYDDDNHCKMEHILGAVRC-----QPLTPN----- 255  
QY 251 KSMPEKFKRLGKMTKEEMMI FIEASSENPCLYSYV--CYSPOAEKVYKLCSPADLHKV 308  
DB 256 -----FLKEQM-----KNCVDLARKVPACREYLA-KIFK-----DLTLHKC 289  
QY 309 GTVV--TPND--IYAG-----QVPLKQTKTNHSTKSLQTAFA-- 344  
DB 290 PGVKERTPTMTMTI FVAGGFRHSLDILBAVNDVDMTWTLLANLRI--PRSLGAAPLK 346  
QY 345 -----RTVNCFY--WPD--AQQNTWPKPTMPLFVRIKPSLVCCGYYAIGSD 388  
DB 347 GKPYAVGGRNNNGSSYDSDDWDRYSAYVETWPCAPMSVPRHRVGVAAMDELMAYVGG- 405  
QY 389 SVGGELNRRRTVERIDTEKDEWTVSPLPACAMQMSAAVVVHDCIYVM-----TLNLKVC 441

DB 406 SAGMEYH-NTVEYIDPDIDRMTLVOPMFAKRLGVVVVNNLLVAIGFPDGNERLASVBC 464  
QY 442 YFPRSDSVWEMAMRQTSRSPASAAAFGDKIFYIGLHATNSGIRLPGTVDGSS--VTV 499  
DB 465 YHPENNEMSFLEPLDTGSGAGVAALNNGYIYVGGF-----DGTQLATV 509  
QY 500 EYDVNKNEMKQANIPAKR 519  
DB 510 ERYDTENDTMDVAPIQIAR 529

Search completed: July 14, 2003, 18:20:53  
Job time : 33.9026 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:10:22 ; Search time 11.4598 Seconds  
(without alignments)  
4345.419 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789  
Sequence: 1 MGSPAPGALGYREFTNR.....RQKGAATNTKVHILGSP 518

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2556	91.6	480	2	JC7812 BCL6 homologous zi
2	1128.5	40.5	706	2	152586 B-cell CLL/lymphom
3	1122.5	40.2	706	2	A48752 B-cell CLL/lymphom
4	559	20.0	794	2	S59069 Z13 protein - mous
5	482	17.3	610	2	JC7315 myonectin - human
6	481	17.2	688	2	A56360 zinc finger protei
7	434.5	15.6	614	2	JH0500 zinc finger protei
8	432.5	15.5	555	2	I53869 zinc finger protei
9	424	15.2	555	2	G02075 transcription repr
10	421.5	15.1	673	2	S36336 probable transcrip
11	419	15.0	676	2	I50643 gammaFib-C - chick
12	417.5	15.0	701	2	T14757 hypothetical prote
13	415	14.9	803	2	S26823 zinc finger protei
14	413.5	14.8	828	2	A32891 zinc finger protei
15	412.5	14.8	693	2	I37570 zinc finger protei
16	408	14.6	728	2	A48830 probable transcrip
17	407	14.6	209	2	S47068 zinc finger protei
18	404	14.5	668	2	T08725 probable finger pr
19	401	14.4	1191	2	S35305 zinc finger protei
20	399	14.3	469	2	I36600 zinc finger protei
21	399	14.3	485	2	A40751 zinc finger protei
22	397	14.2	543	2	B34612 zinc finger protei
23	396	14.2	449	2	S41647 zinc finger 5 prot
24	394.5	14.1	594	2	T12488 hypothetical prote
25	394	14.1	196	2	S06561 zinc finger protei
26	393.5	14.1	466	1	TWPF transcription fact
27	393	14.1	475	2	S03679 zinc finger protei
28	392	14.1	292	2	S43826 zinc finger protei
29	390	14.0	393	2	JN0533 zinc finger protei

30	389.5	14.0	195	2	S00754 zinc finger protei
31	389.5	14.0	733	2	JC7679 dendritic cell-der
32	387.5	13.9	553	2	S22954 zinc finger protei
33	387.5	13.9	675	2	S51037 zinc finger protei
34	387	13.9	325	2	I38616 zinc finger protei
35	387	13.9	448	2	S71427 transcription fact
36	386.5	13.9	580	2	A37107 spermatogenesis pr
37	386	13.8	654	2	A57785 zinc finger protei
38	385	13.8	572	2	I39311 Kruppel-type zinc
39	384.5	13.8	347	2	S00549 developmental cont
40	383.5	13.8	169	2	A39240 zinc finger protei
41	382.5	13.7	540	2	B57785 zinc finger protei
42	381	13.7	201	2	I57505 zinc finger protei
43	380.5	13.6	247	2	S47070 zinc finger protei
44	380.5	13.6	261	2	S70006 zinc finger protei
45	380.5	13.6	710	2	I48668 zinc finger protei

## ALIGNMENTS

## RESULT 1

JC7812 BCL6 homologous zinc finger protein BAZF - human  
C:Species: Homo sapiens (man)  
C>Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 03-Jun-2002  
C:Accession: JC7812  
R:Sakashita, C.; Fukuda, T.; Okabe, S.; Kobayashi, H.; Hirose, S.; Tokuhisa, T.; Miyas  
Biochem. Biophys. Res. Commun. 291, 567-573, 2002  
A>Title: Cloning and characterization of the human BAZF gene, a homologue of the BCL6 on  
A:Reference number: JC7812; PMID:11855826; MIM:21845446  
A:Accession: JC7812  
A:Molecule type: DNA  
A:Residues: 1-480 <SAK>  
A:Cross-references: DDBJ:AB076580; DDBJ:AB076581  
C:Comment: This protein, a zinc finger protein containing a conserved amino terminal BTB  
rly in the induction of megakaryocytic differentiation to produce mature platelets. This  
cell fate.  
C:Genetics:  
A:Gene: bazf  
A:Map position: 17p13.1  
A:Introns: 60/2; 134/2; 256/2; 297/3; 353/1; 396/3; 442/3  
C:Keywords: differentiation

Query Match	Score	DB 2:	Length	DB 1:
Beet Local Similarity	92.7%	Pred. No. 3.5e-137;		
Matches	480;	Conservative	0;	Mismatches 0; Indels 38; Gaps 1;
QY	1	MGSPAPGALGYREFTNRSSDVLGNINELRLRGILTDVTLVGGCPRLRAKAVLIACS	60	
DB	1	MGSPAPGALGYREFTNRSSDVLGNINELRLRGILTDVTLVGGCPRLRAKAVLIACS	60	
QY	61	GFYISIFRGAGVGVVLSLPGGPARGPAPLDMYTSRLSLSPATAPAVLAATYLOM	120	
DB	61	GFYISIFRGAGVGVVLSLPGGPARGPAPLDMYTSRLSLSPATAPAVLAATYLOM	120	
QY	121	EHVQACHRFQASVEPGLISRLPEAPPTPTAPPGSPRRSGHDPPTESCSGQ	180	
DB	121	EHVQACHRFQASVEPGLISRLPEAPPTPTAPPGSPRRSGHDPPTESCSGQ	180	
QY	181	PPSPAPDPKACNMKKRYIVLNSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSS	240	
DB	181	PPSPAPDPKACNMKKRYIVLNSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSS	240	
QY	241	SSSSSEEGPIPGQSRSLPTAATVQFKGAPASTYLLTQAODTSGSPSEARLPQSE	300	
DB	241	SSSSSEEGPIPGQSRSLPTAATVQFKGAPASTYLLTQAODTSGSPSEARLPQSE	300	
QY	301	PPSCNCEAVAGCSGLSLVPGDEDKPKCOLCSSFRYGNLASHRTYHTGKPYHCS	360	
DB	301	PPSCNCEAVAGCSGLSLVPGDEDKPKCOLCSSFRYGNLASHRTYHTGKPYHCS	360	
QY	361	ICGARFNPANLKTSHRISHGEKPKYCTCGSRFVQVRSQPPSGFGKPARGCVQKGF	420	

Db 361 ICARFRPANTKTHSHIGEKYKCTGSRFV----- 395  
Qy 421 CSQGRDLKSPQVAVLRAHVLIHTGEKYPCTGTRPRHIOTLKSHVHTGKPYH 480  
Db 396 -----QVALRRAHVLIHTGEKYPCTGTRPRHIOTLKSHVHTGKPYH 442  
Qy 481 CDEGLHFRHKSQRLRLRQKGAATNTKVYHILGSP 518  
Db 443 CDEGLHFRHKSQRLRLRQKGAATNTKVYHILGSP 480

RESULT 2  
B-cell CLL/lymphoma 5 (BCL5) protein - human  
N/Alternate names: finger protein LAZ-3; zinc finger transcription factor BCL-5  
C/Species: Homo sapiens (man)  
C/Date: 01-Nov-1996 #sequence, revision 01-Nov-1996 #text\_change 31-Dec-2000  
C/Accession: I52586; S40520; S32767  
R/MtL: T.; Kawamata, N.; Hirohata, S.; Aoki, N.  
Blood 83, 26-32, 1994  
A/Title: Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5.  
A/Reference number: I52586; MUID:94100541; PMID:8274740  
A/Accession: I52586  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-706 <MIK>  
R/Character: J.P.; Dewelndt, C.; Tilly, H.; Quidet, S.; Lecocq, G.; Bastard, C.  
Nature Genet. 5, 66-70, 1993  
A/Title: LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3  
A/Reference number: S40520; MUID:94035122; PMID:8220427  
A/Accession: S40520  
A/Molecule type: mRNA  
A/Residues: 1-706 <KER>  
A/Cross-references: EMBL:Z21943; NID:G297025; PIDD:CAA79937.1; PID:G297026  
C/Genetic:  
A/Gene: GDB:BCL5; LAZ-3; ZNF51  
A/Cross-references: GDB:125178; OMIM:151441; GDB:138176; OMIM:109565  
A/Map position: 17q22-17q22; 3q27-3q27  
A/Introns: 60/3  
C/Superfamily: POZ domain homology  
C/Keywords: DNA binding; transcription factor; zinc finger  
F/18-121/Domain: POZ domain homology <POZ>  
F/520-541/Region: zinc finger CCH motif  
F/548-568/Region: zinc finger CCH motif  
F/576-596/Region: zinc finger CCH motif  
F/604-624/Region: zinc finger CCH motif  
F/632-652/Region: zinc finger CCH motif  
F/660-681/Region: zinc finger CCH motif

Query Match 40.5%; Score 1128.5; DB 2; Length 706;  
Best Local Similarity 35.0%; Pred. No. 1.5e-56;  
Matches 270; Conservative 48; Mismatches 117; Indels 337; Gaps 16;

Qy 1 MGSPAPBEGALGVBEFTRHSSDVLGNLRLGLITDVTLLVGGQPLRAKAVLIACS 60  
Db 1 MASPA--DSCI-----QFTRHNSDVLNLNRLSRDILTDVIVSREQFRAKTVLMACS 54  
Qy 61 GFYVSTFRGAGVGVVLSLPGSPBAGRPALDDPWYTSRLRSPATAPVLAATYLLQ 120  
Db 55 GLFYVSTFTDQDKCNLSVINDPEINPEGFCILLDFWYTSRLNREGNIMAVMATYLLQ 114  
Qy 121 EHVVAQCHRFIOAS----- 134  
Db 115 EHVVDCTCRKTKASAEAMVSATKPRBEPLNSRLMLPQDITMAYGRGVENNLLPLASAPG 174  
Qy 135 -----YEPAGIS----- 141  
Db 175 CESRAPAPSLYGLSTPPASYSWYSHLPVSGLLFSDBEFVDVMPVAPPPKRALPCDS 234  
Qy 142 -----LRPL----- 145

Db 235 ARPVGESRPTLEWSPNVCHSNISPKETIBEARSDMHWAVAGLKPAPASARNAPYF 294  
Qy 146 -----EAEPPT-----PPTAP-----PGSPRSSEGGPDPTESRS----- 176  
Db 295 PCDKASKEERFSSDEELHFEPPNAPLNKGLVSPSPQSDQPNPTSCSSKAC 354  
Qy 177 ---CSGPPSPASPDKACNMKKYKYLINSQASQAGSLVGRSSGQCPQARLPBGDEA 233  
Db 355 ILQASGSPPAKSPDPAKCNMKYKYLIN----- 384  
Qy 234 SSSSSSSSSSEEGRIIPQSRSLPTATVQPKGAP-----ASTPYLLTQAQDT- 285  
Db 385 -----SLNQNAKPEGEQALGRLSPPRAYTAPACQPMEPENLDLQSPFKLSASGEDST 439  
Qy 286 -----SGSP-----SERAPPT----- 296  
Db 440 IPQASRLNIVNRMSTGSRSSSESHSPLYMHPKCTSGGSSQPHAEMLHTAGPTPE 499  
Qy 297 ---PGSEF-----FSCQNCBAVAGCSSGLD-SLVPGDEKPYKQQLCRSFRYK 342  
Db 500 EMGETOSEYSDSSCENGAFPCNECDRCFSEASLKRHTLQTHSDRPYKCDRCQASFRYK 559  
Qy 343 NLASHRTVYTGKPYHCSTCGARFNRPNALKTHSHIGEKYKCTGSRFVQVRSQRP 402  
Db 560 NLASHRTVYTGKPYRCNDCGAQFNRPALKTHSHIGEKYKCTGSRFV----- 612  
Qy 403 SGFGKPARGVGQKGFCCSSQRODLKSPQVAVLRAHVLIHTGEKYPCTGTRPRH 462  
Db 613 -----QVALRRAHVLIHTGEKYPCTGTRPRH 641  
Qy 463 LQTLKSHVLIHTGEKYPHCDPCGLHFRHKSQRLRLRQKGAATNTKVYH 514  
Db 642 LQTLKSHVLIHTGEKYPHCKCNLHFRHKSQRLRLRQKGAATNTKVYH 693

RESULT 3  
A48752  
B-cell CLL/lymphoma 6 (BCL6) protein - human  
N/Alternate names: zinc finger transcription factor BCL-6  
C/Species: Homo sapiens (man)  
C/Date: 07-Oct-1994 #sequence, revision 07-Oct-1994 #text\_change 01-Dec-2000  
C/Accession: A48752  
R/Title: B.H.; Lieta, F.; Cocco, F.L.; Knowles, D.M.; Offit, K.; Chaganti, R.S.K.; Dalla-Fav  
Science 262, 747-750, 1993  
A/Title: Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphc  
A/Reference number: A48752; MUID:94053709; PMID:8235596  
A/Accession: A48752  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-706 <YEA>  
A/Cross-references: GB:U00115; NID:G392426; PIDD:AAC50054.1; PID:G392427  
C/Genetic:  
A/Gene: GDB:BCL6; BCL5; LAZ3; ZNF51  
A/Cross-references: GDB:138176; OMIM:109565  
A/Map position: 3q27-3q27  
C/Superfamily: B-cell CLL/lymphoma 5 protein; POZ domain homology  
C/Keywords: transcription factor; zinc finger  
F/18-121/Domain: POZ domain homology <POZ>  
F/520-541/Region: zinc finger CCH motif  
F/548-568/Region: zinc finger CCH motif  
F/576-596/Region: zinc finger CCH motif  
F/604-624/Region: zinc finger CCH motif  
F/632-652/Region: zinc finger CCH motif  
F/660-681/Region: zinc finger CCH motif

Query Match 40.2%; Score 1122.5; DB 2; Length 706;  
Best Local Similarity 36.3%; Pred. No. 3.3e-56;  
Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;

Qy 1 MGSPAPBEGALGVBEFTRHSSDVLGNLRLGLITDVTLLVGGQPLRAKAVLIACS 60  
Db 1 MASPA--DSCI-----QFTRHNSDVLNLNRLSRDILTDVIVSREQFRAKTVLMACS 54

QY 61 GFFYSIFRGRAGVUDVLSLPGGPPKAGAPLIDFMYTSLRLSPATAPAVLAATYLOM 120  
| | | | |  
DB 55 GLFYSIFRQTLKCNLSVINLDEINDEGFCILDFMYTSLRLREGNIMAVATATYLOM 114  
| | | | |  
QY 121 EHVYQACHRFQAS----- 134  
| | | | |  
DB 115 EHVVDTCRKFQKASAEAMVSAIKPPEEPLNSMLMPQDIMAYRGVEVNNILPLRSADG 174  
| | | | |  
QY 135 -----YEPGLGIS----- 141  
| | | | |  
DB 175 CESRAFAPLSYGLSTFPASYSMTSLPVSLLFSDDEFRDVMFVAMPFKERALPCDS 234  
| | | | |  
QY 142 -----LRPL----- 145  
| | | | |  
DB 235 ARPVEEYRPTLEVPNVCHSNISPKETIPEARSDMHVSVAEGLKPAASARAPYF 294  
| | | | |  
QY 146 -----EAPPT-----PPTAP-----PPGPRRSEGHDPPTERS----- 176  
| | | | |  
DB 295 PCDKASKEEERPSSEDEIALHFEPPNAPLNRKGLVSPQSPQSDCQPNPTEACSSKNAC 354  
| | | | |  
QY 177 ---CSGPPSPASPDPKACNMKKYKTYLVNS-----QASQAGSLVGERSSGQPCP 223  
| | | | |  
DB 355 IIOAGSSPPAKSPPTDPKACNMKKYKTYLVNSLQNAKPGPEQAEIGRLSPRAYTPAPAC 414  
| | | | |  
QY 224 QARLPSEGDASSSSSSSSSEEGPIPGPQSL-----SPTAATVQF----- 266  
| | | | |  
DB 415 QPPM--EPENLDLQSPFKLSAGSDSTIP--QASRLNNIYNRSMTGSPRSSSESHPLYMAP 472  
| | | | |  
QY 267 ----KCG--APASTPYLLTSQ----AODTSGSSEARPLPGESEFSSCONCEVAGCSSG 316  
| | | | |  
DB 473 PCTSGSGSPQHAEMCLHTAGPTFAEEMGETOSEYSDSSCENGAFPCNECDRFESEAS 532  
| | | | |  
QY 317 LD-SLVPGDDEKPYKQLCRSSSPRYKGNLASHRTVTGEGYHCISIGARFNRPANLKT 375  
| | | | |  
DB 533 LGRHILQTHSDRKYKCDRCQASGRYKGNLASHKTVTGEGYRCNLCGAQFNRPANLKT 592  
| | | | |  
QY 376 SRHSGEKPYKCECTGSRFVQVRSPPSGFQGRAGVGVQKGFCSQRDLKSPSGCV 435  
| | | | |  
DB 593 TRHSGEKPYKCECTCGARFV----- 614  
| | | | |  
QY 436 AHLRAVLLHTGKPRPCPTCTRRRLQTLKSHVRLHTGKPYHCDPCGLHFRHKSQR 495  
| | | | |  
DB 615 AHLRAVLLHTGKPRPCPTCTRRRLQTLKSHVRLHTGKPYHCDPCGLHFRHKSQR 674  
| | | | |  
QY 496 LHLRQKGAATNTKVHYHI 514  
| | | | |  
DB 675 LHLRQKGAATNTKVHYRV 693  
| | | | |

## RESULT 4

559069  
C1:Species: Mus musculus (house mouse)  
C1:Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999  
C1:Accession: S59069  
R1:Schulz, T.C.; Hopwood, B.; Rathjen, P.D.; Wells, J.R.B.  
Biochem. J. 311, 219-224, 1995  
A1:Title: An unusual arrangement of 13 zinc fingers in the vertebrate gene Z13.  
A1:Reference number: S59069; PMID:96003919; PMID:7575457  
A1:Accession: S59069  
A1:Status: preliminary  
A1:Molecule type: mRNA  
A1:Residues: 1-794 <SCH>  
A1:Cross-references: GB:U14556; NID:g608136; PIDN:AAA5493.1; PID:g608137  
C1:Superfamily: POZ domain homology  
F1:0-108/Domain: POZ domain homology <POZ>

Query Match 20.0%; Score 559; DB 2; Length 794;  
Best Local Similarity 26.6%; Pred. No. 2,2e-24;  
Matches 164; Conservative 73; Mismatches 198; Indels 182; Gaps 17;  
QY 16 EFRHSSDVLGNLNLRLGILTDVTLVGGQPLRAHKAVALIACGFPYIFRGRAG-VG 74  
| | | | |  
DB 62 EFRHSSDVLGNLNLRLGILTDVTLVGGQPLRAHKAVALIACGFPYIFRGRAG-VG 74  
| | | | |

DB 2 DFFQHSQRVLEQNLNQROLGLICDTFVVDGVDFRAHKAVALIACSEYRMLFVDQKDVH 61  
| | | | |  
QY 75 VDVLSPGGEARAGAPLIDFMYTSLRLSPATAPAVLAATYLOMENVQACHRFQAS 134  
| | | | |  
DB 62 LDI-----SNAAGIGQVLEFFYTKLSLSPENVUDVLAASFLQMODIVTACH-LKSL 114  
| | | | |  
QY 135 YEPGLGISRLPPEAPPTPTAPPSSPRRSEGHDPPT----- 172  
| | | | |  
DB 115 AEPSSSTGESADA-----SAVEGDDKRAKDEKAAATMLSLQAGRSSSTGPRELKE 167  
| | | | |  
QY 173 -----ESRSCSQ-----PPSPASDPPACNMKKYKTYLVNSQASQASL----- 212  
| | | | |  
DB 168 ERGGQAEASASGAQETEKADAREPPVLEKDDPTS-----SMAAEARALSBS 216  
| | | | |  
QY 213 -----VGERSSGQPCPQ-----ARLPSEGDASSSSSSSSSEEGPIPGP 253  
| | | | |  
DB 217 SEQEMNEVERPASGEGQEBEGAGPATVKEGHLNLNGEPRENESSAGTDSQGLQMEGQ 276  
| | | | |  
QY 254 QSRLEPTAATVQFK-----CGAPASTPYLLTSQAODTSGSPSEBARPLPGESEFF 302  
| | | | |  
DB 277 NLRSGTGYGDRTESKAYGSIHKCEDCKEFT-----HTGNFKRHRLHTGKPYCT 325  
| | | | |  
QY 303 SCQNGEAV-----AGSSGLDSLVPDEDEKPYKQLCRSSFRY----- 340  
| | | | |  
DB 326 SCRECSKAFSDPAACKAHKTHSP---LKYGCEBCGKSYRLISLNLHKRHSGEARVR 382  
| | | | |  
QY 341 -----KGNLASHRTVTGKPYHCISIG----- 363  
| | | | |  
DB 383 CGDDCGKLPITTSNLRQHLVHSGQPKYQCDYGRFSFSDPTSMRLETHDTDKHKCHPC 442  
| | | | |  
QY 364 -ARFNRPANLKTSHRSHSGEKPYKCECTGSRFVQVRSPPSGFQGRAGVGVQKGFCS 422  
| | | | |  
DB 443 DKFPQVGLKAKHLKTHIADGPLKCECKQGT-----TSGNLKRHLRSHSGEKPYVCT 496  
| | | | |  
QY 423 SGRODLKSPSPVAHLRAVLLHTGKPRPCPTCTRRRLQTLKSHVRLHTGKPYHCD 482  
| | | | |  
DB 497 HCQRFADPGA---LQHVRLHTGKPKQCVICGAFTQASSLIAHVRQHTGKPYCE 552  
| | | | |  
QY 483 PCGLHFRHKSQRRLHLR 499  
| | | | |  
DB 553 RCKGRFVQSSQLANHIR 569  
| | | | |

## RESULT 5

JC7315  
C1:Species: Homo sapiens (man)  
C1:Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 08-Sep-2000  
C1:Accession: JC7315  
R1:Alliel, P.M.; Seddidi, N.; Goudou, D.; Cifuentes-Diaz, C.; Romero, N.; Velasco, E.; Ri  
Biochem. Biophys. Res. Commun. 273, 385-391, 2000  
A1:Title: Myoneurin, a novel member of the BTB/POZ-zinc finger family highly expressed in  
A1:Reference number: JC7315  
A1:Accession: JC7315  
A1:Molecule type: mRNA  
A1:Residues: 1-610 <ALB>  
A1:Cross-references: GB:AF14884  
A1:Comment: This protein, belonging to the family of eukaryotic BTB/POZ and zinc finger p  
and a regulator in the genital tract.  
C1:Keywords: tandem repeat; transcription regulation; zinc finger

Query Match 17.3%; Score 482; DB 2; Length 610;  
Best Local Similarity 26.8%; Pred. No. 3.7e-20;  
Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;  
QY 16 EFRHSSDVLGNLNLRLGILTDVTLVGGQPLRAHKAVALIACGFPYIFRGRAG--GV 73  
| | | | |  
DB 2 QVSHHCHEHLRLNKRQBRAGFLDCCTIVIGEFQFAHNVLASFSYFGAIVRSTSENNV 61  
| | | | |  
QY 74 GVDVLSLQGPARGAPLIDFMYTSLRLSPATAPAVLAATYLOMENVQACH----- 128  
| | | | |  
DB 62 FLDQSQV---KADFQKLEBITYGTINLDSMWVKEIHQAADVLYK/BEVVYTKCKIKMED 117  
| | | | |

```

Qy 129 -----RFIQASYEPLGISLRP-----LEAPRPPTAPPPG 159
Db 118 FAFIAPSSSTEISSITGINELNOQTCLLRDYNNEKSEVSTDLIQANQKQALAKSS 177
Qy 160 -SPRSEGHDPPTESRSCSGP-----PSPASBPACAKWK---KY 197
Db 178 QTKKKKAFNPSPTGQNKTVQYPSDILLENASVELFLDANLPPPVVEQVAQINDSELEL 237
Qy 198 KYVLVNSQASQ--AGSLVGRSSSGGPCPARLPS---GDEASSSSSSSSSEEGPIRP 252
Db 238 TSVENTFPADIVHTVTVKRGKSGOPNCLAKHKSMSNLASYSKSPLEANSSEE---LDO 295
Qy 253 POSRLSPATATVQFKCAPASTPYLLTSQAQDTSGSPSEBARP---LPSEFFSCQNC 308
Db 296 RYSAKAPMCNT-----CGKVSEBASLRRHRIHKG-----VKPYCHLCGKAFQCNQK 346
Qy 309 AVAGCGSGLDLVPGEDEKPYQCQLCRSSFRYKGNLASHRTVTTG--EKPHGICGARR 367
Db 347 TVHRTHTG-----EKPYKCELCDFKQACQOLVFHSHRMHGEKPYKCDVCMQFA 397
Qy 368 RPAHLKTHSRHISGEKPYKCEGSRPVQVRSQPPSGFGKPARAGVGQKGFCSGQRD 427
Db 398 TSSNLKTHAKHSGEKPYPVDRCGQRF-----424
Qy 428 LKSPSQVAHLRAHVLHTGEKPYPCPTGTRFRHLQTLKSHVRIHTEKPYHCDPCGLH 487
Db 425 -----AQASTLVTVHVRHTEKPYVCDTGKAFVASSLSLTHSRKHTEKPYICGICGKS 479
Qy 488 FRHKSQRLHLRQKHG 503
Db 480 FISSGELNKHFRSHTG 495

```

## RESULT 6

```

A56360
zinc finger protein (clone 18) - human
N/Alternate names: GLI-related finger protein HKR3
C/Species: Homo sapiens (man)
C/Date: 03-Oct-1995 #sequence_reviation 03-Oct-1995 #text_change 05-Nov-1999
C/Accession: A56360; E31201
R/Sugawara, M.; Scholl, T.; Ponath, P.D.; Strominger, J.L.
Mol. Cell. Biol. 14, 8438-8450, 1994
A/Title: A factor that regulates the class II major histocompatibility complex gene DPA
A/Reference number: A56360; MUID:95059073; PMID:7969177
A/Accession: A56360
A/Molecule type: mRNA
A/Residues: 1-688 <SUG>
A/Cross-references: GB:L16896; NID:9292934; PIDN:AAA65124.1; PID:9292935
R/Ruppert, J.M.; Kinzler, K.W.; Wong, A.C.; Bigner, S.H.; Kao, F.T.; Law, M.L.; Senanez,
Mol. Cell. Biol. 8, 3104-3113, 1988
A/Title: The GLI-Kruppel family of human genes.
A/Reference number: A93103; MUID:89096896; PMID:2850480
A/Accession: E31201
A/Molecule type: DNA
A/Residues: 461-516 <RUP>
C/Genetics:
A/Introns: 506/1
A/Note: the list of introns may be incomplete
C/Keywords: DNA binding; zinc finger
F/12-111/Domain: POZ domain homology <POZ>

```

Query Match 17.2%; Score 481; DB 2; Length 688;

Best Local Similarity 27.7%; Pred. No. 4.8e-20;

Matches 160; Conservative 72; Mismatches 216; Indels 130; Gaps 22;

```

Qy 17 FTRHSDVIGNLNLRGLITVTLVGGQPLRAHKAVALIACGFFYSIFRAGAVGV 76
Db 5 FVGHSTRVVLDELNKKQREKQYCDATLDVGLVFKAHMSVLACSHFQSYLGGSGGSV 63
Qy 77 VLSLPGPEKRGAPALIDFMYTSRLRSPATPAVALAATYLOMEHNVQACHFT 131
Db 64 --VLPAQ-FAEIRGLLDFYTGHLATSGNDQVLLAAELRVPAVELCQSFKPTSV 120

```

```

Qy 132 -QAS-----YEPILG-----ISLRPLEAP--PTPTAPPPGSP 161
Db 121 GQAGGQSGSLGPASQNNVSHYKEPAGLEEEVSTLGLVPRDQPRGSHSPQRQLHSP 180
Qy 162 RSEGHDPPT-----SRSC-----QGPSAPAPDRACMKKYYKTYL-- 202
Db 181 AQSEG-----PSSLCKLKOALPKCSLEDKKPEDCVPPPLPABGAOLQGSNEMEVVQ 236
Qy 203 -----NSQASQAGSLVGRSSG---OPCPQARLPSGDEASSSSSSSSSEEGPIRP 253
Db 237 VEDDGDGTSWSEPEAVLTRKSNVIRKPC-----AAPALASAGLAEE-----P 280
Qy 254 QSRLSPTATVQFKCAPA-----STPYLLTSQAQDTSGSPSEBARLPSEFFSCQNC 307
Db 281 AENRGTAVPVE-----CPTCHKKFLSKYLLKVNHRKHTEKPE--CPKCGKCYFRKENTL 334
Qy 308 --EAVAGCGSGLDLVPGEDEKPYQCQLCRSSFRYKGNLASHRTVHTGEKPHGICGAR 365
Db 335 LEHEARNCMNRSEQVLP-----CSVQCEFRRRMELRVHVSHTGEMPYKSCSSCQO 386
Qy 366 FNRPAHLKTHS--RHISGEKPYKCEGSRPV--QVRSQPPSGFGKPARAGVGQKGFCSG 423
Db 387 FMQCKDLOSMTKLHGAPRPACTCAKCFLSRTLOLHEAFKHGEKLFVCEGCHRAS 446
Qy 424 QKODLK-----SP-----SQVAHLRAHVLHTGEKPYPCPTGTRFRHLQ 464
Db 447 SRNGLOMHIKAKHNRBRPVHCFCSHAFTQKANLMLHRLTHGKPYFCQHLGKTFRTQA 506
Qy 465 TLKSHVRIHTEKPYHPCDGLHFRHKSQRLHLRQKH 502
Db 507 SLDKNRTHTGERPFSCEQRFTEKGPLLHVAASH 544

```

## RESULT 7

```

JH0500
zinc finger protein 29 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1992 #sequence_reviation 31-Mar-1992 #text_change 08-Dec-2000
C/Accession: JH0500
R/Denny, P.; Ashworth, A.
Gene 106, 221-227, 1991
A/Title: A zinc finger protein-encoding gene expressed in the post-meiotic phase of sper
A/Reference number: JH0500; MUID:92039080; PMID:1937051
A/Accession: JH0500
A/Molecule type: mRNA
A/Residues: 1-614 <DEN>
A/Cross-references: GB:X55126; NID:955470; PIDN:CAA38920.1; PID:955471
A/Experimental source: testis
C/Keywords: DNA binding; zinc finger
F/217-244/Region: zinc finger
F/245-272/Region: zinc finger
F/273-300/Region: zinc finger
F/301-328/Region: zinc finger
F/329-356/Region: zinc finger
F/357-384/Region: zinc finger
F/385-412/Region: zinc finger
F/413-440/Region: zinc finger
F/441-468/Region: zinc finger
F/469-496/Region: zinc finger
F/497-524/Region: zinc finger
F/525-552/Region: zinc finger
F/553-580/Region: zinc finger
F/581-608/Region: zinc finger

```

Query Match 15.6%; Score 434.5; DB 2; Length 614;

Best Local Similarity 30.3%; Pred. No. 1.8e-17;

Matches 122; Conservative 42; Mismatches 133; Indels 105; Gaps 15;

```

Qy 162 RSEGHDPPTESRSCSGSPSPASP--DPKAKMKKYYTYITLNSQASQAGSLVGRSS- 218
Db 170 QQSDGDSQD---FERDCGSGGAGGHAPEGDPF-----VPSBGRVGLIGLQGT 216

```

QY 219 -CG---PCFOA-----RLPG-----DEASSSSSSSSSEBGPDPGQ 254  
 DB 217 LGEKPEPCQCKTFSRSHLITHERHTGKEYKDECCGKSFSDSNFSRHQTHTGK 276  
 QY 255 SLSLPTAAVQPK---CGAPAS--TPYLLLSQAODTSGSPSERA-----R 294  
 DB 277 -----PYKCRDCGKSFSSRSANLITHRIHTGKEKFPQCECKSFSSPNLIAMOR 326  
 QY 295 PLPGSEFFSCQNCENAVAGSSGIDS---LVPGDEDPYKCOLCRSSFRKYNLASHRTVH 351  
 DB 327 THTGKPYSCPECGKSFGRSSSLNTHOGIHTG--EKPYACKEGSEFSTNSLIRQRH 384  
 QY 352 TGEKPYHCSICGARFRRPNALKTHSHISGEKPYKCTG3SRFVQVRSOPSGFGKPAR 411  
 DB 385 TGEKPYKCTEGCGKSFQSSALITHRTHTGKEKPYQCEGCKNF----- 427  
 QY 412 GGVGQKGFCSQRODLKSP-----SQVAHLRAVLIHTGKEKPYPCPTGTRR 461  
 DB 428 ---SSNSLNATHRRKTHLVKPKYKGLCGKSFQSSSLIAQGTHTGKEKPECLTGSEFS 484  
 QY 462 HLQTKSHVRIHTGKEKPYHCDPCGLHFRHKSQRLRLRQXG 503  
 DB 485 WSNLILKHRTHTGKEKPYRCGDCGKGFQSRQSLVHQRTHTG 526

## RESULT 8

153869  
 zinc finger protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 01-Dec-2000  
 C:Accession: I53869  
 R:Brady, J.P.; Platiogoraky, J.  
 Gene 149, 299-304, 1994  
 A:Title: A mouse cDNA encoding a protein with zinc-fingers and a KRAB domain shows siml  
 A:Reference number: I53869; MUID:95047492; PMID:7959006  
 A:Accession: I53869  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-555 <RES>  
 A:Cross-references: GB:I28167; NID:9758660; PIDN:AAA67545.1; PID:9758661  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.5%; Score 432.5; DB 2; Length 555;  
 Best Local Similarity 28.6%; Pred. No. 2.1e-17;  
 Matches 135; Conservative 52; Mismatches 178; Indels 107; Gaps 18;

QY 79 SLPGSEBAGPRLDPMYTSRLSPATAPAVLAATTIQEHVYQACHRFIOASYEPL 138  
 DB 112 SYVNTPTQLQKAP-----CQSRPRSPAPVSDENYIK-SHTISV----- 149  
 QY 139 GISLRPLEAPPTPTAPPSS-----PRSEGHDP-----PP 171  
 DB 150 -----PIHGRPPLPRLAQSTSLTRQTRKQOYHOGSHFDEAGQKRSHTSWLPRR 204  
 QY 172 TESRSC-SQGPSPSPAS---PDPKACMKYKYIVLNSQAEQAGSLVGERSSGQPCPARL 227  
 DB 205 SDHRCVCKSGRCSPDHGRDGVVPRFDQNGGIHTQKXSQRHEATDLDSSSD--PHQL 262  
 QY 228 PGDEDESSSSSSSSS-----SSEGRPLPGCQSLSPALAVQKCGAPAST 274  
 DB 263 DSGGRACACVECGRPGCTGDCIHOGGRTGEBRKEFGHCRVCTPKPYTCEOC-----E 317  
 QY 275 PYLLTSGAODTSGSPSERARPLPGSEFFSCQNCENAVAGSSGIDS---LVPGDEDPYK 331  
 DB 318 KALLVDQHLMSHVKVTTRRP-----YKERTGSAFSAQSHLODQRLHTG--EKPYK 369  
 QY 332 QLCRSSFRYKGNLASHRTVHTGKEKPYHCSICGARFRRPNALKTHSHISGEKPYKCTG 391  
 DB 370 DACGKSFSSSHLRSHQVHTGKEKPYKCGCGKSFCSNLTLYHQHVTHTGKEKPYKCVDCG 429  
 QY 392 SRFVQVRSOPSGFGKPARGV--GQXGFGS--SQRODLKSPSQVAHLRAVLIHTG 447  
 DB 430 KEFSR-----PSLQ---AHQGIHTGKSYVCTMCGKGYTLNS-----NLQVHLRVHTG 475

QY 448 EKPYPCPTGTRFRHLQTLKSHVRIHTGKEKPYHCDPCGLHFRHKSQRLRLHR 499  
 DB 476 EKPYSCDVQKGFSSRSQSLQSHQVHTGKEKPYKQDVCKSGKSGMRNLIIHR 527

## RESULT 9

G02075  
 transcription repressor zinc finger protein 85 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 01-Dec-2000  
 C:Accession: G02075  
 R:Poncelet, D.A.  
 submitted to the EMBL Data Library, September 1995  
 A:Reference number: G09169  
 A:Accession: G02075  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-595 <PON>  
 A:Cross-references: EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g1017722  
 C:Genetic: GDB:ZNF85  
 A:Gene: GDB:ZNF85  
 A:Cross-references: GDB:132279  
 A:Map position: 19p12-19p12  
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.2%; Score 424; DB 2; Length 595;  
 Best Local Similarity 38.5%; Pred. No. 6.7e-17;  
 Matches 90; Conservative 26; Mismatches 92; Indels 26; Gaps 5;

QY 267 KCGAPASTPYLLTSGAODTSGSPSERARPLPGSEFFSCQNCENAVAGSSGTL--DSLVP 323  
 DB 206 ECG-----KAFWSSSTLTGKHILHTGKEKPYKCECGKAFNQSSNLIKHKIHTG 254  
 QY 324 DEDKPYKCOLCRSSFRYKGNLASHRTVHTGKEKPYHCSICGARFRRPNALKTHSHISGEK 383  
 DB 255 --EKPYKCECGKAFNRSTLTTHKIHTGKEKPYKCGKAFNRSTLTTHRKIHTGK 312  
 QY 384 PYKCEGSRFVQVRSOPSGFGKPARGVGQKGFCSQRODLKSPSQVAHLRAVLIHTG 443  
 DB 313 PYKCECGKAFNQSSNLTHKI-----HTGKPYKCKCKGAP---NSAHILTHEV 362  
 QY 444 IHTGKPYPCPTGTRFRHLQTLKSHVRIHTGKEKPYHCDPCGLHFRHKSQRLRLH 497  
 DB 363 IHTGKPYKCECGKAFNHSHTLTTHKIHTGKEKPYKCKCEGKAFNHSHTLTTH 416

## RESULT 10

S36336  
 probable transcription factor PLZF - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 05-Nov-1999  
 C:Accession: S36336; S31989  
 R:Chen, Z.; Brand, N.J.; Chen, A.; Chen, S.J.; Tong, J.H.; Wang, Z.Y.; Waxman, S.; Zeller  
 EMO J. 12, 1161-1167, 1993  
 A:Title: Fusion between a novel Krueppel-like zinc finger gene and the retinoic acid rec  
 A:Reference number: S36336; MUID:93209216; PMID:8384553  
 A:Accession: S36336  
 A:Molecule type: mRNA  
 A:Residues: 1-673 <CHB>  
 A:Cross-references: EMBL:Z19002; NID:g38517; PIDN:CAA79489.1; PID:g38518  
 C:Genetic:  
 A:Gene: PLZF  
 C:Superfamily: POZ domain homology  
 C:Keywords: zinc finger  
 F/20-118/Domain: POZ domain homology <POZ>

Query Match 15.1%; Score 421.5; DB 2; Length 673;  
 Best Local Similarity 25.3%; Pred. No. 1.1e-16;  
 Matches 155; Conservative 63; Mismatches 236; Indels 159; Gaps 19;  
 QY 20 HSDVILGNLNLRLGLITDVTLLVGGPLRAHKAVALIACSGFYFYSIF-RGAGVGVNVL 78

Db 16 HPTGLCKANQMLAGTLCVIMVDSQEFHARTVLACTSKMFEILFHRSQHTYLLDFL 75  
QY 79 SLPGREARGFARLDPMYTSRLSLSPATAPAVLAATYQMEHVQACGRF---TQASV 135  
Db 76 S-----PKTFQOILEYATYATLQAKAEDLDLLYAEILEIYLEIBQCKLMETIQASD 129  
QY 136 E----- 136  
Db 130 DNDTEATMADGAAEBEEDRKARYLKNIFISKSSSESGVASVAGSLPGMVDQSPSVST 189  
QY 137 -----PLGISRLPLEAEPPPTAPPPGSPRRSEGHDPPTESRSCS 178  
Db 190 SFGLSAMSPYKPAVDSIMTIGOSILQGTLOPPAGPEEPPLAGGRRPGVAEVXTMMQYD 249  
QY 179 QGPPSPAPDPKACNMKKYKYIVLNSQASQAGSLVGRSSGQCPQAR-----LPS 229  
Db 250 E---VPSQDSFGAAE-----SLISGAGKQVEBKGEKGTPTRSSVITSARELHY 297  
QY 230 GDEASS-----SSSSSSSSSEEGPIPGPSRLSPATATVQFKGAPASTPYLLTS--- 280  
Db 298 GRESABQVPRPAPADAGARTGRRENHAPREPKHLGISVLPNKHADAVLSMPSVTSGLH 357  
QY 281 --QAQTSQSPSERAPRLPGS---EPFSCQNCBAVAGSSGGLSLVPGDEDEKPYKQQLC 334  
Db 358 VQPALAVSMDFSYGGILLPGQFLQRELFIS-KLGEVLAVGMS--ESRTTIGQ-----CSYC 409  
QY 335 RSSEFRYKGNLASHTVYHTEKPHCSICGARFNRPNLKTSHRHS-GEKPYKCEGCGSR 393  
Db 410 GVLPDNEAVEQIRKHLHSGMKYTGCELCGRFLDSLRLNHLHLAHGAKAFCDQCGAG 469  
QY 394 F-----VQVRSQPPSG-----FQGRPA-----RGVGQKGGFCSSQRODLKS 430  
Db 470 FSKEDALETHRQHTHTGDMAVFCLLCGRKFOAQSAHQHMEVHAGVRSYICSECNRTF-- 527  
QY 431 PRQVAVHLRAVHLIRHGEKPYPCPTGCTRRHLQTLKSHRITHTGEKPYHCDPCGHLFHR 490  
Db 528 -PSHTA-LKRLHLSHGDHPRECEFCGSCFRDSTLKSHRITHTGEKPYECNCDDKFSL 585  
QY 491 KSQRLHLRLROKHG 503  
Db 586 KQOLETHRYVHTG 598

RESULT 11  
150643  
gammaRFP-C - chicken (fragment)  
C/Species: Gallus gallus (chicken)  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 01-Dec-2000  
C/Accession: 150643; 150641; 150642; S44242; S44264; S44265  
R/LIU, Q.; Shalaby, F.; Puri, M.C.; Tang, S.; Breitman, M.L.  
Dev. Biol. 165, 165-177, 1994  
A/Title: Novel zinc finger proteins that interact with the mouse gamma F-crystallin prom  
A/Reference number: 150641; PMID:94374555; PMID:8088434  
A/Accession: 150643  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-676 <LIU>  
A/Cross-references: EMBL:X79011; NID:g475903; PIDN:CAAS5644.1; PID:g475904  
A/Accession: 150641  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 'MRVHRELGMWA', 23, 'GSGR', 28-676 <LI2>  
A/Cross-references: EMBL:X79051; NID:g479055; PIDN:CAAS5653.1; PID:g479056  
A/Accession: 150642  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 36-676 <LI3>  
A/Cross-references: EMBL:X79050; NID:g479057; PIDN:CAAS5652.1; PID:g479058  
C/Superfamily: POZ domain homology  
F:49-152/Domain: POZ domain homology <POZ>

Query Match 15.0%; Score 419; DB 2; Length 676;  
Best Local Similarity 23.7%; Pred. No. 1.5e-16;

Matches 146; Conservative 67; Mismatches 184; Indels 219; Gaps 21;  
QY 9 GALTGYREPTRESSVLTGNLNLRLRGITLDTVTLVGGQPLRAHRAVLJACGPFYSIR 68  
Db 34 GMLBAMEVPSRQQLQLNTQRTKGFCDVITVQNALFRAHKOITLAASAYLSLV- 92  
QY 69 GRAGVVDVSLPGREARG-FAPLLDFMYTSRL-RLSPA---TAPAVLAATYQMEHV 123  
Db 93 ----VHNDLNLNDHEVNSPGIFRLILDFITYRLDECEPCEGSGVAGVLAASYSYQIPQL 148  
QY 124 VQACNR-----FIQASYE-PLGISRLPLE----- 146  
Db 149 VALCKKQLKRSKGYCHLRGAPRYKLGRGLRATTPVIAQVSGTRPVDLQVEBAAPLN 208  
QY 147 -----A 147  
Db 209 TCGGELVASASGTPLHPHGLCPREHSCSPGGLDLKSKSPGPSAQLLPTRLRLPAER 268  
QY 148 EPTTPP-----TAPPGSPRRSEGH---DPPTESRSCSQGPSP 184  
Db 269 EPLPFRHDSPPVSGGLAHGPAAYKDSPPGEF---GSHPATDPPSTPPCAE-PLPL 324  
QY 185 ASPPDPKAC-NMKKYK---YIVLSQASQAGSLVGRSSGQF--CPQARLPG--GDEASS 235  
Db 325 RQDGRLEMYRMKKHPELPGYL---DEGAKEKLEEREKAESEPPAAPQPPYPSVESNDLEP 381  
QY 236 SSSSSSSSSSEEGPIPG-----POSRLSPATATVQFKGAPASTPYLLTSQA 282  
Db 382 DNSTSEBTSSSGEPFGDALDHYCNHLYEPBSLDNLVYCIPLCGKFPSSQLAHAYEA 441  
QY 283 QDTs-----GSPERARPLGSEPFSCNCEAVAGSSGGLSLVPGDEDEKPYKQQLCRSF 338  
Db 442 HNEEELVHKAALAEQVLPF-----LDKGAAGL-----GDILRPYKSSCDKSY 483  
QY 339 RYKGNLASHTVYHTEKPHCSICGARFNRPNLKTSHRHSIHGEKPYKCEGCGSRFVQVR 398  
Db 484 KDPATLRQHEKTHMLTRPYPCITCGKFTQRTWTRHMRSHGLRPFACDAGMRF---- 539  
QY 399 SQPPSGQKPARGGVQGGFCSSQRODLKSPQVAVHLRAVHLIRHGEKPYPCPTGCT 458  
Db 540 -----TRQRLTBRHMRHSGEKPYECQVCG 565

QY 459 RFRHLQTLKSHVRIH 473  
Db 566 KFAQQRNLISHMKM 580

RESULT 12  
T14757  
hypothetical protein DKFZp572C163.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 01-Dec-2000  
C/Accession: T14757  
R/Wamput, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A/Reference number: Z18181  
A/Accession: T14757  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-701 <WAM>  
A/Cross-references: EMBL:AL110217  
C/Genetics: A/Experimental source: adult subthalamic nucleus; clone DKFZp572C163  
A/Note: DKFZp572C163.1  
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.0%; Score 417.5; DB 2; Length 701;  
Best Local Similarity 41.6%; Pred. No. 1.8e-16;  
Matches 89; Conservative 23; Mismatches 77; Indels 25; Gaps 5;

QY 298 GSEFFSCNCEAVAGCCSGL---DSLVPDEDEKPYKQQLCRSSFRYKGNLASHTVHGE 354  
Db 352 GKPYECNECEKPTFAHNSALRAHQNHTG--EKLVECSGCGKTFQKTRLSTHRIHGE 409



QY 355 KEYHCSIGCARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQVRSPSGFOGKPA---- 410  
DB 410 KEYECSKCKGTFKSKSYLTSCHERHITGSEKPYECNCGKTFVY-----KALILYH 458  
QY 411 -RGCGVQKQKGFSSORODLKSPPSOVAHLRAHVLITGSEKPYPCPTGTRFRLIQLTKSH 469  
DB 459 QIHITGSEKPYECNCGKTFVY-----SQRTHLCAHQRHITGSEKPYECNCGKTFADNSALRAH 514  
QY 470 VRIHTGSEKPYPCDPCGLAFHRHKSQRLHLRQKHG 503  
DB 515 HRIHTGSEKPYECNCGKTFKSHLRAHLRTRSG 548

## RESULT 13

zinc finger protein ZNF43 - human  
S26823  
N/Alternate names: zinc finger protein kor27  
C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000  
C/Accession: S26823; 137967; S10416  
R/Lovering, R.; Trowsdale, J.  
Nucleic Acids Res. 19, 2921-2928, 1991  
A/Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell 11  
A/Reference number: S26823; MUID:91279444; PMID:1711675  
A/Accession: S26823  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-803 <LOV-  
A/Cross-references: EMBL:X59244; NID:938031; PIDN:CAA11932.1; PID:938032  
R/Thiesen, H.J.  
New Biol. 2, 363-374, 1990  
A/Title: Multiple genes encoding zinc finger domains are expressed in human T cells.  
A/Reference number: 137949; MUID:91145339; PMID:2288909  
A/Accession: 137967  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 476-531 <THI>  
A/Cross-references: EMBL:X52358; NID:934160; PIDN:CAA36584.1; PID:930090  
C/Genetics:  
A/Gene: GDB:ZNF43; HTP6  
A/Cross-references: GDB:128653  
A/Map position: 19p13.1-19p12  
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C/Keywords: DNA binding; zinc finger

Query Match 14.8%; Score 415; DB 2; Length 803;  
Best Local Similarity 30.2%; Pred. No. 2.9e-16;  
Matches 91; Conservative 34; Mismatches 118; Indels 58; Gaps 6;  
QY 267 KCGAPASTYVLLTSAQDNTSGSP-----SERARPLPGSEFPSCQNC-E 308  
DB 312 ECGKAFNWSTLTAKHRIHTGSEKPYCEBGCAFNQFSLTTKRIHTLAKFYKTECEB 371  
QY 309 AVAGCSGGLDSLVPGDEKPYKQLCRSSFRYKGNLASHRTVHTGSEKPYHCSIGCARFNR 368  
DB 372 AVSRSSNLTKKKHITGSEKPYKCEBGCAFKMSSKLTETKLTHTGSEKPYKCEBGCAFNW 431  
QY 369 PANLTKHSHRHSIGSEKPYKCECTGSRFVQ-----RSQPPSGFOGKPARGVGQK 418  
DB 432 PSLTGTGRIHTGSEKPYKCEBGCAFNQFSLTTKRIHTLAKFYKTECEB 489  
QY 419 GFCSSORODLKSPPSOV-----AHLRAHVLITGSEKPYPCPTGTRFRLIQLTKS 468  
DB 490 NLTKKKHITGSEKPYKCEBGCAFKMSSKLTETKLTHTGSEKPYKCEBGCAFNW 549  
QY 469 HVRITGSEKPYHCDPCGLAFHRHKSQRLHLR-----OKHGAATNTVY 510  
DB 550 HKRIHTGSEKPYKCEBGCAFTQSSNLTKKKHITGSEKPYKCEBGCAFTQSSNLTKKKI 609  
QY 511 H 511  
DB 610 H 610

## RESULT 14

finger protein 1, placental - human  
A32891  
C/Species: Homo sapiens (man)  
C/Date: 20-Oct-1989 #sequence\_revision 20-Oct-1989 #text\_change 01-Dec-2000  
C/Accession: A32891  
R/Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncellet, D.A.; Belayew, A.; Martial, J.  
DNA 8, 377-387, 1989  
A/Title: The human genome contains hundreds of genes coding for finger proteins of the  
A/Reference number: A32891; MUID:89377476; PMID:2505992  
A/Accession: A32891  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-428 <BEL>  
A/Cross-references: GB:M27877  
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C/Keywords: tandem repeat

Query Match 14.8%; Score 413.5; DB 2; Length 428;  
Best Local Similarity 42.0%; Pred. No. 1.9e-16;  
Matches 87; Conservative 24; Mismatches 77; Indels 19; Gaps 5;

QY 302 FSCQNCBAVAGCSGLDS--LVPGDEKPYKQLCRSSFRYKGNLASHRTVHTGSEKPYH 358  
DB 33 FKCDICGKI FNKKSNLASHORHTG--EKPYKCEBGKVFHNMSHLAQHRRIHTGSEKPYK 90  
QY 359 CSICGARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQVRSPSGFOGKPARGVGQK 418  
DB 91 CNEGKVFNQSLHQAQRHITGSEKPYKCEBGKVFHNI-----SHLAQHRRIHTGSEKPY 144  
QY 419 GFCSSORODLKSPPSOVAHLRAHVLITGSEKPYPCPTGTRFRLIQLTKSHRITGSEKPY 478  
DB 145 YECNCKGKVF-----SRNSYLVQHLITGSEKPYRCNCGKVFHNSHLAQRHITGSEKPY 200  
QY 479 YHCDPCGLAFHRHKSQRL-----RLHLRQK 501  
DB 201 YKCNCGKVFHNSKSLVNWHRITGSEK 227

## RESULT 15

zinc finger protein - human (fragment)  
I37570  
C/Species: Homo sapiens (man)  
C/Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 01-Dec-2000  
C/Accession: I37570  
R/Abirink, M.; Aveskog, M.; Hellman, L.  
DNA Cell Biol. 14, 125-136, 1995  
A/Title: Isolation of cDNA clones for 42 different Kruppel-related zinc finger proteins  
A/Reference number: I37566; MUID:95169271; PMID:7865130  
A/Accession: I37570  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-693 <RES>  
A/Cross-references: EMBL:X78927; NID:9498726; PIDN:CAA5527.1; PID:9498727  
A/Genetics:  
A/Gene: HZF4  
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 14.8%; Score 412.5; DB 2; Length 693;  
Best Local Similarity 41.4%; Pred. No. 3.5e-16;  
Matches 87; Conservative 18; Mismatches 76; Indels 29; Gaps 4;

QY 298 GSEFSCQNCBAVAGCSGLD-SLVPGDEKPYKQLCRSSFRYKGNLASHRTVHTGSEKPY 356  
DB 419 GEKPYKCEBGCAFNQFSLTTKRIHTGSEKPYKCEBGCAFNQFSLTTKRIHTGSEKPY 478  
QY 357 YHCSIGCARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQ-----VRSQPPSGFOG 407  
DB 479 YKCEBGCAFNRAIDLKRIHTGSEKPYKCEBGCAFNQFSLTTKRIHTGSEKPYKCEBGCAFN 530  
QY 408 KPARGVGQKGFSSORODLKSPPSOVAHLRAHVLITGSEKPYPCPTGTRFRLIQLTK 467

Db	.531	-----GKPKC----	EECGKSFSSAHLQAHQKVTGKPKYKCGECGKGFKWSLND	579
Qy	468	SHVRIHTEKPYHCDPGLHFRHKSOLRLH	497	
Db	580	MHQRVHTGKPYTCGACGKHFSQASSLQLH	609	

Search completed: July 14, 2003, 18:23:16  
Job time : 14.6598 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 5.38782 Seconds

(without alignments)  
3987.653 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789

Sequence: 1 MGSPAPEGALGYVREFTRH.....ROKGAATVTKVHILGSP 518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1138	40.8	707	1 BCL6_MOUSE	P41183 mus musculu
2	1128.5	40.5	706	1 BCL6_MOUSE	P41182 homo sapien
3	575.5	20.6	803	1 Z151_HUMAN	Q13105 homo sapien
4	558	20.0	794	1 Z151_MOUSE	Q60821 mus musculu
5	482.5	17.3	697	1 Y441_HUMAN	Q43167 homo sapien
6	476	17.1	688	1 HKR3_HUMAN	P10074 homo sapien
7	434.5	15.6	614	1 ZF28_MOUSE	Q07330 mus musculu
8	427.5	15.3	645	1 ZF93_MOUSE	Q61116 mus musculu
9	424	15.2	595	1 ZN85_HUMAN	Q16600 homo sapien
10	421.5	15.1	458	1 Z239_HUMAN	Q05516 homo sapien
11	421.5	15.1	673	1 Z145_HUMAN	Q14526 homo sapien
12	420.5	15.1	733	1 HIC1_HUMAN	Q96135 homo sapien
13	419	15.0	615	1 HIC2_HUMAN	Q96135 homo sapien
14	419	15.0	676	1 HIC1_CHICK	Q90850 gallus gall
15	418	15.0	892	1 HIC1_MOUSE	Q91715 mus musculu
16	415.5	14.9	803	1 Z226_HUMAN	Q9NYT6 homo sapien
17	415	14.9	476	1 Y106_HUMAN	Q96319 homo sapien
18	415	14.9	803	1 ZN43_HUMAN	P17038 homo sapien
19	413.5	14.8	407	1 OZF_MOUSE	Q62513 mus musculu
20	413.5	14.8	407	1 ZN83_HUMAN	P51522 homo sapien
21	412.5	14.8	698	1 Z234_HUMAN	Q14588 homo sapien
22	412	14.8	597	1 ZF37_HUMAN	Q9Y6G3 homo sapien
23	411.5	14.8	818	1 KR18_HUMAN	Q9HCG1 homo sapien
24	410.5	14.7	606	1 Z214_HUMAN	Q9U159 homo sapien
25	409.5	14.6	616	1 ZN93_HUMAN	P35789 homo sapien
26	407	14.6	209	1 ZF93_HUMAN	Q14590 homo sapien
27	406.5	14.6	574	1 YE73_HUMAN	Q9P255 homo sapien
28	406.5	14.6	629	1 Z195_HUMAN	Q14628 homo sapien
29	404.5	14.5	446	1 ZN38_HUMAN	P17036 homo sapien
30	404	14.5	734	1 ZN42_HUMAN	P28698 homo sapien
31	403	14.4	683	1 Y972_HUMAN	Q9Y2H8 homo sapien
32	402.5	14.4	504	1 Z205_HUMAN	Q95201 homo sapien
33	402.5	14.4	682	1 ZN45_HUMAN	Q02386 homo sapien

34	402	14.4	407	1 OZF_MOUSE	Q62981 rattus norv
35	401	14.4	706	1 Z151_CHICK	Q90625 gallus gall
36	401	14.4	1191	1 ZN91_HUMAN	Q05481 homo sapien
37	400	14.3	741	1 Z288_HUMAN	Q9HC78 homo sapien
38	399.5	14.3	913	1 Z228_HUMAN	Q9UJ33 homo sapien
39	399	14.3	469	1 Z135_HUMAN	P52742 homo sapien
40	398	14.3	489	1 E21T_HUMAN	Q9NQZ8 homo sapien
41	397	14.2	543	1 ZN08_HUMAN	P17098 homo sapien
42	396.5	14.2	498	1 Z271_HUMAN	Q14591 homo sapien
43	396	14.2	449	1 ZF91_HUMAN	Q43829 homo sapien
44	396	14.2	449	1 ZF91_MOUSE	Q08376 mus musculu
45	396	14.2	604	1 Z300_HUMAN	Q96199 homo sapien

## ALIGNMENTS

```

RESULT 1
ID BCL6_MOUSE STANDARD; PRT; 707 AA.
AC P41183; 061065;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE B-cell lymphoma 6 protein homolog.
GN BCL6 OR BCL-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Skeletal muscle;
RX MEDLINE=96038894; PubMed=7478591;
RA Fukuda T., Miki T., Yoshida T., Hatanano M., Ohashi K., Hirosewa S.,
RA Tokuhisa T.;
RT "The murine BCL6 gene is induced in activated lymphocytes as an
RT immediate early gene.";
RL Oncogene 11:1657-1663 (1995).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=96247530; PubMed=8652841;
RA Allman D., Jain A., Dent A., Maile R.R., Selvaggi T., Kenny M.R.,
RA Staedt L.M.;
RT "BCL-6 expression during B-cell activation.";
RL Blood 87:5257-5268 (1996).
CC - FUNCTION: TRANSCRIPTION REGULATOR THAT PROBABLY PLAY AN IMPORTANT
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - SIMILARITY: CONTAINS 1 BTF/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D38377; BAA07456.1; -
CC EMBL; U41465; AAB17432.1; -
CC HSSP; P15822; 1BBO.
CC MGD; MGI:107187; Bcl6.
CC InterPro; IPR000210; BTF_POZ.
CC InterPro; IPR000822; ZNF_C2H2.
CC Pfam; PF00096; ZF_C2H2; 6.
CC PRINTS; PR00048; ZINCINGER.
CC SMART; SM00325; BTF; 1.
CC SMART; SM00355; ZNF_C2H2; 6.
CC PROSITE; PS50097; BTF; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.

```

DR PROSITE, PSS0157; ZINC FINGER C2H2\_2; 6.  
 KM Nuclear protein; Transcription regulation; Activator; DNA-binding;  
 KW Zinc-finger; Metal-binding; Repeat.

FT DOMAIN 32 99 BTR.  
 FT DOMAIN 519 682 ZINC FINGERS.  
 FT ZN\_FING 519 542 C2H2-TYPE.  
 FT ZN\_FING 547 569 C2H2-TYPE.  
 FT ZN\_FING 575 597 C2H2-TYPE.  
 FT ZN\_FING 603 625 C2H2-TYPE.  
 FT ZN\_FING 631 653 C2H2-TYPE.  
 FT ZN\_FING 659 682 C2H2-TYPE.  
 FT CONFLICT 456 456 A -> G (IN REF. 2).  
 SQ SEQUENCE 707 AA; 78981 MW; 2051DD808D32D5EC CRC64;

Query Match 40.84; Score 1138; DB 1; Length 707;  
 Best Local Similarity 36.44; Prod. No. 3.1e-51;  
 Matches 263; Conservative 51; Mismatches 149; Indels 270; Gaps 14;

QY 1 MGSPAPREGALGVYRETRHSDVGLNLRGLITDVTLLVGGQPLRAKAVLIACS 60  
 1 MASPA--DSCI-----QTRHSDVLLNLRSLRDITDVIVSRHQRAHKTIVLMACS 54  
 DB 61 GFRYSIFRGAGVGVVLSLPGGPEARGFAPLIDFWTSLRLSPATAPVLAATYLOM 120  
 55 GLFYSLFTDQKCNLSYINLDPRISEPGFCLIDFWYTSRLNLRGINAVMTTAYLOM 114  
 QY 121 EHYVQCHRFIOASYEPLGISLRPLAE----- 148  
 115 EHYVDCKRKTKASSEMAMALRPREEFLNRLMHPHIMAYRGREVENNMLRNTRG 174  
 DB 149 -----PPT----- 151  
 175 CESRAFAPPLYGLSTPPAPYMYSHLPSTFLFSDBELDARMPVAPFPERALPCD 234  
 QY 152 ----- 151  
 235 SARQVNEYSRPAMEVSPSLCHSNISPKAEVBEARSIDHVSVPKGPAPVSAANAPY 294  
 152 -----PPTAP-----PGSPRSEGHDPPTESRS----- 176  
 295 FPCDKASKERERSSSEDEILHPEPPAPLNRKGLVSPSPKSDQCPNPTSCSSKQA 354  
 QY 177 -----CSQGPSPAPSPPKACNMKKYKIVLNS-----QASQASLVGRSSGQPCQA 225  
 355 CILQASGSPPAKSPTPDKACNMKKYKIVLNSLQNAKPGSSQALGRLSPRAYAPAP 414  
 DB 226 RLPSGDEAS--SSSSSSSSSSSEGPFGPSRL-----SPTAATVQF----- 266  
 415 CQPMPEPANDLOSPTLSASGSDSTIP--QASRLNMLVNRSLAGSPRSSSHSPMYAP 473  
 267 -KCGASAPFPYLLTQAODTSGSPSEARAPLPQSEF-----PSCQCEAVAGCSSG 316  
 474 PKTSCGSGSPQPTBENCMTHTAGTTPBEMETQSEISDSCENGCTFPCNEDCRSEBS 533  
 DB 317 LD-SLVGDEDEKPKQCLCRSSFRYKGNLSHRTVHTGKPYKCSICGAFNRPANLKT 375  
 534 LKHTLTQTHSDKRYKCDRCQASFRYKGNLSHRTVHTGKPYKCNICGAFNRPANLKT 593  
 QY 376 SRHSEKPKYKCTGSRFVQVNSQPSGCGKPARAGVGQKGFSSQRQDLKSPSQY 435  
 594 TRHSEKPKYKCTGSRFV----- 615  
 DB 436 AHRARAVLHTGKPYPCPTGTRPHLOTLKSHVRIHTEGKPYHCDPGLHFRHSQAR 495  
 616 AHRARAVLHTGKPYPCPTGTRPHLOTLKSHVRIHTEGKPYHCDPGLHFRHSQAR 675  
 QY 496 LHLRQKGAATNTKYVHI 514  
 676 LHLRQKGAATNTKYQVRI 694  
 DB

RESULT 2  
 BCL6\_HUMAN

ID BCL6\_HUMAN STANDARD; PRT; 706 AA.  
 AC P41B2;  
 DT 01-FEB-1995 (rel. 31, Created)  
 DT 01-FEB-1995 (rel. 31, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE B-cell lymphoma 6 protein (BCL-6) (Zinc finger protein 51) (LAZ-3 protein) (BCL-5).  
 GN BCL6 OR ZNF51 OR LAZ3 OR BCL5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NC NCB1 Taxid=9606;  
 RX TISSUE=Skletal muscle;  
 RX MEDLINE=94035122; PubMed=8220427;  
 RA Kerchaert J.-P., Deweindt C., Tilly H., Quief S., Lecocq G., Bastard C.  
 RT "LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas."  
 RL Nat. Genet. 5:66-70(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94053709; PubMed=8235596;  
 RA Ye B.H., Lista F., Lo Coco F., Knowles D.M., Offit K., Chaganti R.S.K., Dalla-Favera R.  
 RT "Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma."  
 RL Science 262:747-750(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Liver;  
 RX MEDLINE=94100541; PubMed=8274740;  
 RA Miki T., Kawamata N., Hirosewa S., Aoki N.  
 RT "Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5, encodes a Kruppel-like zinc-finger protein."  
 RL Blood 83:26-32(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93281738; PubMed=8506375;  
 RA Barton B.W., Nucifora G., McCabe N., Espinosa R. III, le Beau M.M., McKelthan T.W.  
 RT "Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5262-5266(1993).  
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT PROBABLY PLAYS AN IMPORTANT ROLE IN LYMPHOMAGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- DISEASE: INVOLVED IN A FORM OF B-CELL NON-HODGKIN LYMPHOMA CHARACTERIZED BY CHROMOSOMAL TRANSLOCATION T(3;14)(Q27;Q32) AND T(3;22)(Q27;Q11) THAT INVOLVES BCL6 AND IMMUNOGLOBULIN GENE REGIONS.  
 CC -1- DISEASE: INVOLVED IN A FORM OF B-CELL LEUKEMIA CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(3;11)(Q27;Q23) THAT INVOLVES BCL6 AND POU2AF1 (OBF1).  
 CC -1- DISEASE: INVOLVED IN A T(3;4)(Q27;P11) CHROMOSOMAL TRANSLOCATION WITH ARH (TTP).  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromancer/Genes/BCL6ID20.htm".  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC EMBL; Z21943; CAA79937.1; -;  
 CC EMBL; U00115; AAC50054.1; -;  
 CC EMBL; S67779; -; NOT\_ANNOTATED\_CDS.



DR	EMBL; AL034555; CAB85445.1; -	
DR	EMBL; Y09723; CAAT0889.1; -	
DR	EMBL; M88369; AAA61327.1; -	
DR	HSSSP; P07248; IARE.	
DR	TRANSFAC; T03414; -	
DR	GeneW; HGNC:12936; ZNF151.	
DR	MIM; 604084; -	
DR	InterPro; IPR000210; BTB_POZ.	
DR	InterPro; IPR000822; Znf_C2H2.	
DR	Pfam; PF00096; ZF-C2H2; 13.	
DR	Pfam; PF00651; BTB; 1.	
DR	PRINTS; PR00048; ZINC_FINGER.	
DR	ProDom; PD000003; Znf_C2H2; 2.	
DR	SMART; SM00225; BTB; 1.	
DR	SMART; SM00355; Znf_C2H2; 13.	
DR	PROSITE; PS50097; BTB; 1.	
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.	
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.	
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding; Nuclear protein; Repeat.	
FT	DOMAIN	1 104
FT	BTB.	
FT	DOMAIN	1 104
FT	ZN_FINGERS.	
FT	ZN_FING	306 328
FT	ZN_FING	334 356
FT	ZN_FING	362 384
FT	ZN_FING	390 412
FT	ZN_FING	418 440
FT	ZN_FING	446 468
FT	ZN_FING	474 496
FT	ZN_FING	502 524
FT	ZN_FING	530 552
FT	ZN_FING	558 580
FT	ZN_FING	586 608
FT	ZN_FING	614 637
FT	ZN_FING	739
FT	CONFLECT	73
FT	SEQUENCE	803 AA; 87959 MW; FEB7BBA4E14AE200 CRC64; M -> V (IN REF. 2).

Query Match	20.6%	Score 575.5;	DB 1;	Length 803;
Best Local Similarity	29.4%;	Pred. No. 1.2e-22;		
Matches 180;	Conservative 64;	Mismatches 205;	Indels 163;	Gaps 20;

```

Oy 16 EFTRRSSDVLGNLNLRLRGITDVTLLVGGORPLANKAVILACGFFSIIRGAG--VG 74
Db 2 DFGQISNHLBGLMORQGLGLCDCTPVDVDYGNFPAKRAVLACGEYFMFLFVDQKYVH 61
Oy 75 VDVLISLPGGEARGAFLPLDEWYTSRLRLSPRTABAVLAATAVYLOMENVQACHRFQAS 134
Db 62 LDI-----SNAAGGOMLEFMYTAKLSTISPNVDVLAVALFLOMODITACH----- 109
Oy 135 YEPGLISLRLAEAPPTR-----TAPPPGPRRSECHPDPPTSSCSQSPSPAPPD 188
Db 110 -----ALKSLAEPAISPGNAELALTBGGKRAKEEKVATSTLSRLQARSTPIGP- 161
Oy 189 PKACMKYKYKIVILNSOASOAGSLGERSSSGCPCEQARPL--SGDEASSSSSSSSSSS 245
Db 162 --SRDLKEBGGQAGSAAAGAQTEKADAPRPPVELKDPDPTSGMAALAAALASESE 219
Oy 246 EBGPL-PPQPSLSTLAATVPQKCA-PAST---PYLLTSCA-----QDTSGSPS-- 290
Db 220 QEMEYEPARKGEBOEQEOBEOBEEBAGPAVEBEGSOLENGEAPENENEBEAGTDSQOE 279
Oy 291 --ERARPLP-----GSEF-----PSCQNC 307
Db 280 LGSERGLRSGYGERTESKAYGSIYHKCEDCGKFTYTGANKFRIRLHTEKPPSCREK 339
Oy 308 EAV-----AGSSGSLDSLVPGEDKPKYQOLCNSSFRY----- 340
Db 340 SKAFEDPAPACKAHEKTHSP--LKEYGECEBCKSKYRLISLNLHKHSGSGEARVRCEDCG 396
Oy 341 -----KGNLAHRTYHTEKEKPHGSIICG-----AFPN 367
Db 397 KLFTTSGULKRRQVLVHSEKPRQCDYCGARSPSDPTSKMRLHETDTEKHKCPHCDKFN 456

```

Oy 368 RPA<sup>NT</sup>LTHSR<sup>II</sup>HSGEKPY<sup>CT</sup>CGSR<sup>FV</sup>QVNS<sup>OP</sup>SGFG<sup>QK</sup>PA<sup>GV</sup>Q<sup>KG</sup>GFSS<sup>OR</sup>OD 427  
 Db 457 QVNL<sup>LA</sup>HLK<sup>II</sup>HADE<sup>PL</sup>CKRC<sup>GQ</sup>FT-----TSG<sup>LN</sup>KLK<sup>HL</sup>HS<sup>II</sup>SGE<sup>KPY</sup>C<sup>II</sup>HC<sup>Q</sup>RQ 510  
 Oy 428 LKSP<sup>QV</sup>ANL<sup>RA</sup>HA<sup>VL</sup>HTGEKPY<sup>CP</sup>TCG<sup>TR</sup>FR<sup>HL</sup>Q<sup>TL</sup>KS<sup>HR</sup>IT<sup>HS</sup>GEKPY<sup>HC</sup>PC<sup>GL</sup> 487  
 Db 511 FADPG<sup>A</sup>----LQ<sup>NR</sup>HA<sup>II</sup>HTGEKPY<sup>CC</sup>VM<sup>CG</sup>ALFT<sup>Q</sup>ASS<sup>LI</sup>HA<sup>Q</sup>HTGEKPY<sup>CC</sup>RC<sup>GR</sup> 566  
 Oy 488 FR<sup>HS</sup>Q<sup>LR</sup>L<sup>HL</sup>R 499  
 Db 567 FV<sup>Q</sup>SSQ<sup>LA</sup>NH<sup>IR</sup> 578

## RESULT 4

ID	Z151 MOUSE	STANDARD;	PRT;	794 AA.
----	------------	-----------	------	---------

DT 15-JUL-1998 (Rel. 36, Created)

DT 16-OCT-2001 (Rel. 40, Last annotation update

DE protein) (LP-1) (Zinc finger protein 213).

**Mus musculus** (Mou)

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

RN [1]  
PD SCIENCE FROM N T

RA Rapp L., Carmichael G.G.;  
Submitted (APR 1986) to the FBI /DOJ Bank/NDPT database

RN [2]  
PD SENTENCE FROM N 2

```
RC STRAIN=CBA; TISSUE=Kidney;
BY MEDLINE=85003818; PubMed=7575457.
```

RA Schulz T.C., Hopwood B., Rathjen P.D., Wells J.R.E.,  
 DT The unusual arrangement of 13 rings in the "C"

RT	213.7	211-210	224 (100%)
RT	213.7	211-210	224 (100%)

CC -I- FUNCTION: MAY FUNCTION AS A HOUSEKEEPING DNA-BINDING PROTEIN THAT  
CC REGULATES THE EXPRESSION OF SPECIFIC GENES

cc -|- SUBCELLULAR LOCATION: Nuclear (potential).

CC EXAMINED.  
CC 1. SIMILARITY. REFUSAL TO THE WITNESSES' FAULTY OR GOOD FINDING

CC FINGER PROTEINS.  
CC SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN

-----  
 CC The SWISS-DBOT survey is produced through a collaboration  
 CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC The European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

cc articles requires a license agreement (see <http://www.1bd-bid.cn/announce/> or send an email to [1bcbid@163.com](mailto:1bcbid@163.com))

CC -----  
 DB -----  
 EMRI: IT2396: AAAA64848 1: -

DR EMBL: U14556; AAA85493.1; -.  
HSSP: P08046: 1A1H  
DR

DR MGD; MG1:10/410; ZEP100:  
DB InterPro: IPR000310: BTB POZ

DR Incepio; 1PR000822; ZH1\_C2H2.  
DR Pfam: PF00096: zf-C2H2: T3

DR PLAIN; PF00651; BIB; 1.  
DR PRINTS: PR00048: ZINCETINGER

```

DR      PRODOM; PD0000003; ZHL_C2H2; 1.
DB      SMAPT; SM00225; PTR; 1

```

DR SMART; SM00355; ZNF\_C2H2; 13.  
DR PROSITE: PS50097: BTB: 1

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 13

transcription regulation; DNA-binding; zinc-finger; metal-binding; KM

```

KW Nuclear protein; Repeat.
FT DOMAIN 1 104 BTB.
FT ZINC 297 730 ZINC FINGERS.
FT ZN_FING 297 319 C2H2-TYPE.
FT ZN_FING 325 347 C2H2-TYPE.
FT ZN_FING 353 375 C2H2-TYPE.
FT ZN_FING 381 403 C2H2-TYPE.
FT ZN_FING 409 431 C2H2-TYPE.
FT ZN_FING 437 459 C2H2-TYPE.
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 493 515 C2H2-TYPE.
FT ZN_FING 519 543 C2H2-TYPE.
FT ZN_FING 549 571 C2H2-TYPE.
FT ZN_FING 577 599 C2H2-TYPE.
FT ZN_FING 605 628 C2H2-TYPE.
FT ZN_FING 708 730 C2H2-TYPE.
FT ZN_FING 507 529 G -> A (IN REF. 2).
FT CONFLICT 573 573 N -> K (IN REF. 2).
SQ SEQUENCE 794 AA; 86664 MW; FF88565DEB7ED CRC64;

Query Match
Best Local Similarity 20.0%; Score 558; DB 1; Length 794;
Matches 164; Conservative 72; Mismatches 199; Indels 182; Gaps 17;

QY 16 EFTRHSDVYGNLNLRLGILTVTLVGGQPLRAHKA VLIACSGFFYSIFRGRAG-VG 74
DB 2 DEFGHSQVLEQDINQORQLGLCDCTFVVGVDFKAKHKA VLAACSEYFKMLFVDQDVH 61
QY 75 VVVLISPGPEARGAPLIDFWYTSRLRSPATAPAVLAATYLOMEHVOACHRTQAS 134
DB 62 LDI-----SNAGLGQVLEFWYATLSTSPENVDVLA VASFLQMDIVTCHT-LKSL 114
QY 135 YEPPLGISLRLAEPPPTAPPSPRSRSEGHDPPT----- 172
DB 115 AEPSSITGSSAA-----SAVEGDKAKADEKAATMISRLGARGSSITGPELKE 167
QY 173 ---BSRSCSQG-----PPSPAPDPKAKCNMKYKYIVLNSQASQAGSL--- 212
DB 168 ERGGQAESASGAQETEKADAPREPPELKPPTS-----SMAAEALSSBS 216
QY 213 ---VGRSSSGQPPQ-----ARLPSGDENSSSSSSSSSEEGTIPRP 253
DB 217 SSGMEVEPASKGDEQEEGAGPATVKEGMLDGEPRHEEESAGTDSGQELMEQO 276
QY 254 GSRLEPTATVQFK-----CGAPASFPYLLTSQMDTSGSPSRAPLPSESEF 302
DB 277 NLRSGYGRTESKAVGSIHKCEDGCKEPT-----HTGNFKHRIHTEKEFP 325
QY 303 SCONCEAV---AGCSSGLDSLVPDEDPKPKQOLCRSSPRY----- 340
DB 326 GCREGSKAFSDPAACAHKHTSP---LKPYGCEEGSKYRLISLNLKHHSGBARVR 382
QY 341 -----KGNLASHRTVHTGEKPYHCSICG----- 363
DB 383 CGDCGKLTFTSGNLKQHQLVHSGQKYOCDYCGRSFSDPTSKMRHLEHTDTEHKCPAC 442
QY 364 -ARPNRAPLKTSHRHSGEKPKCECTGSRFVQVNSOPPSGFGKPARGVGQKGFSS 422
DB 443 DKFPNOVGHLKAKHLKHLIDGPKCKECCGKQFT-----TSGNLKHLIHSGEKPYVCL 496
QY 423 SGRODLKSPSOVAHLRAHVLHTGKPYPCPTGTRFRHLQTLKSHVHHTGEKPYHCD 482
DB 497 HGORGFADP-----GGQRHVIRIHTGKPCQCVICGAFTQASLIHVNQHTGEKPYVCE 552
QY 483 PCGHLFRHKSQRLHNR 499
DB 553 RCGKRPVGSQLANHNR 569

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIA00441.
GN KIA00441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=brain;
RC MEDLINE=98116655; PubMed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro."
RL DNA Res. 4:307-313 (1997).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB007901; BAA23713.1; -.
DR HSSP; P08046; 1AIG.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000210; BTB POZ.
DR InterPro; IPR000822; ZnF C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF02178; AT hook; 1.
DR PRINTS; PR00048; ZINC FINGER.
DR ProDom; PD000003; znf C2H2; 2.
DR SMART; SM00384; AT hook; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; ZnF C2H2; 8.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS50028; ZINC FINGER C2H2_1; 8.
DR PROSITE; PS50157; ZINC FINGER C2H2_2; 8.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 10 133 BTB.
FT ZINC 294 512 ZINC FINGERS.
FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 322 344 C2H2-TYPE.
FT ZN_FING 350 372 C2H2-TYPE.
FT ZN_FING 378 400 C2H2-TYPE.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT ZN_FING 462 484 C2H2-TYPE.
FT ZN_FING 490 512 C2H2-TYPE.
SQ SEQUENCE 697 AA; 78292 MW; F2BD33C144626544 CRC64;

Query Match
Best Local Similarity 17.3%; Score 482.5; DB 1; Length 697;
Matches 156; Conservative 74; Mismatches 225; Indels 133; Gaps 19;

QY 6 APEGALGYRREFTRHSDVYGNLNLRLGILTVTLVGGQPLRAHKA VLIACSGFFYS 65
DB 5 SPSPGQLVHSDAIDSDVLAASFQDRKKKGFCDITLIVENVHPRDHLALAAASSEYFEM 64
QY 66 IFRGAGVGVLDVLSIPGPEARGAPLIDFWYTSRLRSPATAPAVLAATYLOMEHVO 125

```

Db 65 MFAEBEIGQSIYMLB-GWVADTFGLLEFIYLYGLHASEKSTQIILATQAFUKVYDLVK 123  
 Qy 126 ACHRFIOASYEPGLISLRLEAPPTPTAPPP--GSPRSEGHPPPTESRECSGPPS 183  
 Db 124 AYIDFQNNHSP-----KPTLTNAGAVVVIYINKN--DPPKRRKQ----- 163  
 Qy 184 PASPDPKACNMWKYK-----YIYINSAQASQSLVGERSSQCPQOARLPBGDEA 233  
 Db 164 ----RKKVNTLQEEKSELAAEBEIDLRVYNSVQNFVVKDGSGLVNEQI---AAKER 216  
 Qy 234 SSSSSSSSSSSSEGP1-----PGQSLSPLPA--ATYQFCGAPASPTLYL 278  
 Db 217 ESEPPFCEPSREEMVEKENDYDPKTEGQASQSRYSKRRIRWSYVLK-----DYKL 269  
 Qy 279 TSOADPTSGSPSE--DARPLPGSEFPSCNCEAV-----ACSSGSLDVLPGDD 326  
 Db 270 VGDQED-HGSAKKICGRKRRKPGPEARCCDGVFKYNNPLAHQSHTG-----E 319  
 Qy 327 KPYKCOLCRSSFFRYKNLASHRTVHTGEKPYHCSIGAFENRPAULKTHSRHSGEKPYK 386  
 Db 320 RPKKNECGKGFAGKSLQVHTMHTGERPYCTCVCSKALTTGSLLEHMSLHSGKSPFT 379  
 Qy 387 CETGSRFPQVNRQPPSGFGKPARAGVGQKGFCSQRODLKSPSQVAHLRAVYLIT 446  
 Db 380 CDQCGRYFSQNR-QLKSHYR-----VHTGSHLPECKDCHRKEMDVSQLKGLRTHT 429  
 Qy 447 GEKPYCPGCTGFRHLQTLKSHVRLHTGEKPYHCDPCGLHF-----RH----- 490  
 Db 430 GKRPFTCEICGSKFTAKSSLQTHIRIRHGEKPYSCGICGKSPSDSAAKRRHCLHTGKPR 489  
 Qy 491 -----KSQLRLRLROKGA-----ATNTKVHYHL 515  
 Db 490 FSCPECNLQFARLDNKAHLKHSKEMASDASISGSSNTEVERNIL 537

RESULT 6  
 HKR3 HUMAN STANDARD; PRT; 688 AA.  
 ID AC P10074;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Kruepel-related zinc finger protein 3 (HKR3 protein).  
 GN HKR3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95059073; PubMed=7969177;  
 RA Sugawara M., Scholl T., Ponath P.D., Strominger J.L.,  
 "A factor that regulates the class II major histocompatibility  
 complex gene DPA is a member of a subfamily of zinc finger proteins  
 that includes a Drosophila developmental control protein.";  
 RT Mol. Cell. Biol. 14:8438-8450 (1994).  
 RL (2)  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=98177483; PubMed=9516840;  
 RA Maris J.M., Jensen J., Sulman E.P., Beltlinger C.P., Allen C.,  
 Biegel J.A., Brodeur G.M., White P.S.,  
 "Human Kruppel-related 3 (HKR3): a candidate for the 1p36  
 neuroblastoma tumour suppressor gene?";  
 RT Eur. J. Cancer 33:1991-1996 (1997).  
 RN (3)  
 RP SEQUENCE OF 461-488 FROM N.A.  
 RA MEDLINE=89096896; PubMed=2850480;  
 RA Ruppert J.M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,  
 Law M.L., Suarez H.N., O'Brien S.J., Vogelstein B.,  
 "The Gli-Kruppel family of human genes.";  
 RT Mol. Cell. Biol. 8:3104-3113 (1988).  
 CC -1- FUNCTION: BINDS TO AND REGULATES THE J AND/OR S ELEMENTS IN MHC II  
 CC PROMOTER.

CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC  
 CC EMBL; I16896; AAA65124.1; -  
 CC EMBL; U45325; AAB08973.1; -  
 CC EMBL; U45324; AAB08973.1; JOINED.  
 CC EMBL; M20677; AAA35989.1; -  
 CC PIR; E31201; E31201.  
 CC HSSP; P03001; 1TF3.  
 CC TRANSFAC; T04971; -  
 CC Genew; HGNC:4930; HKR3.  
 CC MIM: 165270; -  
 CC  
 CC InterPro; IPR000210; BTB\_POZ.  
 CC InterPro; IPR000822; ZnF\_C2H2.  
 CC Pfam; PF00096; ZF\_C2H2\_11.  
 CC Pfam; PF00651; BTB; 1.  
 CC PRINTS; PR00048; ZINC\_FINGER.  
 CC Prodom; PD000003; ZnF\_C2H2; 1.  
 CC SMART; SM00225; BTB; 1.  
 CC SMART; SM00355; ZnF\_C2H2; 11.  
 CC DR PROSITE; PS50097; BTB; 1.  
 CC DR PROSITE; PS50028; ZINC\_FINGER\_C2H2\_1; 9.  
 CC DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 11.  
 CC Transcription regulation: Activator; Zinc-finger; DNA-binding;  
 CC Repeat; Metal-binding; Nuclear protein.  
 CC KW  
 CC FT DOMAIN 26  
 CC FT 291 600  
 CC FT ZN\_FING 291 313  
 CC FT ZN\_FING 319 342  
 CC FT ZN\_FING 350 372  
 CC FT ZN\_FING 378 401  
 CC FT ZN\_FING 407 430  
 CC FT ZN\_FING 436 459  
 CC FT ZN\_FING 465 487  
 CC FT ZN\_FING 493 515  
 CC FT ZN\_FING 521 544  
 CC FT ZN\_FING 550 572  
 CC FT ZN\_FING 578 600  
 CC FT CONFLICT 201 201  
 CC FT CONFLICT 244 244  
 CC FT CONFLICT 350 351  
 CC FT CONFLICT 607 607  
 CC FT SEQUENCE 688 AA, 77054 MW, 17054 MW; EBECCB3D6CBDB524 CR664;  
 CC  
 CC Query Match 17.1%; Score 476; DB 1; Length 688;  
 CC Best Local Similarity 27.1%; Pred. No. 1,1e-17;  
 CC Matches 157; Conservative 73; Mismatches 216; Indels 134; Gaps 22;  
 CC  
 CC 17 FTRHSSDVLGNELRLKGLITDVLVGGQGLRAHKAVALICSGFPFISFGKRGVGD 76  
 CC 5 FQOHSVRYLQELNKKREKQYCDATLDVGLVFKAWSWLACSHFPQSLVYDGGSGSV- 63  
 CC 77 VLSLGGPARGAFPLDPMYTSRLSPATPAVLAATVYQMHVVOACGRFI----- 131  
 CC 64 --VLPAG-FAELFGLLDFYTGHLALNSGRDQVLLARELRVEANVELCSFKPKTSV 120  
 CC 132 -QAS-----YEPDG-----ISLRLEAP--PTPTAPPGSP 161  
 CC 121 GQAAGGQSLGPPASQNVNSHVKEPAGLEBEVSRITGLVPRDQPRGSHSPQRQLHSP 180  
 CC 162 RASGEG-----HDPPTESR--SCSGSPSPSPDPKACNMWKYKITYL----- 202  
 CC 181 AQSEGPSSLCGLKQALKRCPLEDKKPBDC-KVPPRPLEAGALQGGSEWEVVVQVED 239



QY 203 ---NSGASGALVGRSSG---QCPQAPALPSCGEASSSSSSSSSSEGPPIPGQSR 256  
DB 240 DDDGDVMSPEAVLTKRSKSNVIRKPC-----AARPALASGLABE-----PARN 283  
QY 257 LSPFATATVQPKGAPR-----STPYLLTSQAODTSGSSSEARPLPGSEFFSCONCEAV 310  
DB 284 RKGTAVPVE---CPTCHKKFLSKYLLKYNKRKHTEKE-----FECPCC--- 324  
QY 311 AGCSGSLDVLVPGD-----EDKPYKCOLCRSSPFYKKNLASHRTVHTGKPYHCSIGS 363  
DB 325 GCCTYRKKEKLEHEARNCNRSEQVTCGVCQETFRRLHLYHVMVSHTEMPYKCSGS 384  
QY 364 AEFNRPANLKTSH-RHSGEKPYKCECTGSRFY-QVRSQPPSGFGKPARGVGQKGFPC 421  
DB 385 QQFMQKDLQSHMKILHGAAPKAPACTCAKCFLSRTELQHEAFKIRGSKLFTVCECGHR 444  
QY 422 SSGQRODLK-----SPP-----SOVAHLRAHYLHTGKPYPCPTCGTRFRH 462  
DB 445 ASSRGLQWIKAKHNERPHVCEFCSSHAFTQKAMINMTLRHTGKPPQCHLCKTFR 504  
QY 463 LQTLKSHVRIHTGKPYHPCDGLPHRHSQRLHLRQHG 502  
DB 505 QASLDKHNTHTGERPFSCFCEQRTKXGPLLRHVASHN 544

RESULT 7  
ZF29 MOUSE STANDARD, PRT, 614 AA.  
ID ZF29 MOUSE  
AC 007230;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Zinc finger protein 29 (Zfp-29).  
GN ZFP29 OR ZFP-29.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA MEDLINE=92039080; PubMed=1937051;  
RX Denny P., Ashworth A.;  
RT "A zinc finger protein-encoding gene expressed in the post-meiotic  
RT phase of spermatogenesis.";  
RL Gene 106:221-227(1991).  
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION DURING  
CC THE POST-MEIOTIC STAGES OF SPERMATOGENESIS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: IN THE ADULT, PREDOMINANTLY FOUND IN  
CC SPERMATIDS. ALSO PRESENT IN THE EMBRYO.  
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -1- SIMILARITY: CONTAINS 1 SCAN BOX.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL: X55126; CAA8920.1; -.  
DB HSBP; P08047; ISP2.  
DB MGD; MGI:99176; Zfp29.  
DB InterPro; IPR003309; Treg SCAN.  
DB InterPro; IPR000822; Znf\_C2H2.  
DB Pfam; PF00096; zf-C2H2; 14.  
DB Pfam; PF02023; SCAN; 1.  
DB PRINTS; PR00046; ZINCFINGER.  
DB ProDom; PD000003; Znf\_C2H2; 13.

DR SMART; SM00431; LER; 1.  
DR SMART; SM00355; Znf\_C2H2; 14.  
DR PROSITE; PS00804; SCAN BOX; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 14.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 14.  
KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat;  
KW Developmental protein; Transcription regulation; Spermatogenesis.  
FT DOMAIN 69 127 SCAN BOX.  
FT DOMAIN 222 608 ZINC FINGERS.  
FT ZN\_FING 222 244 C2H2-TYPE.  
FT ZN\_FING 250 272 C2H2-TYPE.  
FT ZN\_FING 278 300 C2H2-TYPE.  
FT ZN\_FING 306 328 C2H2-TYPE.  
FT ZN\_FING 334 356 C2H2-TYPE.  
FT ZN\_FING 362 384 C2H2-TYPE.  
FT ZN\_FING 390 412 C2H2-TYPE.  
FT ZN\_FING 418 440 C2H2-TYPE.  
FT ZN\_FING 446 468 C2H2-TYPE.  
FT ZN\_FING 474 496 C2H2-TYPE.  
FT ZN\_FING 502 524 C2H2-TYPE.  
FT ZN\_FING 530 552 C2H2-TYPE.  
FT ZN\_FING 558 580 C2H2-TYPE.  
FT ZN\_FING 586 608 C2H2-TYPE.  
SQ SEQUENCE 614 AA; 68714 MW; 738156A0A9413DF7 CRC64;

Query Match 15.6%; Score 434.5; DB 1; Length 614;  
Best Local Similarity 30.3%; Pred. No. 1,3e-15;  
Matches 122; Conservative 42; Mismatches 133; Indels 105; Gaps 15;

QY 162 RRSRGHPDPPTESRSQSGPPSPASP--DPKAMKKYKYYLVLSNQAQSLVGRSS- 218  
DB 170 QOSDDSD---FERDCSGGAQGHAPGEDPR-----VVPSEGRVQLIGLQGY 216  
QY 219 -GQ---PCGPA-----RLPSG-----DEASSSSSSSSSSSEGIIPRQ 254  
DB 217 LGEKPYECPQCKTFRSRKSHLITHERTYGKYYKDCGKSFDSNFSRHQTTHTBK 276  
QY 255 SRLSPFATVQFK---CGAPAS-TPYLLTSQAODTSGSPSERA-----R 294  
DB 277 -----PYKCDCKGKFSRSANLITTHQRIHTGKRPQACBGKSFSPNLIANOR 326  
QY 295 PLPGSEFFSCONCEAVACSSGLDS--LVPGDEDKPYKCOLCRSSPFYKKNLASHRTVH 351  
DB 327 THTGKPYSCPECGKSFGRSSSLNTHQIHNG--EKPYACEGCGSFYSNVLIRHQRH 384  
QY 352 TGEKPYHCSIGARPNRANLKTSHRHSGEKPYKCECTGSRFYQVRSQPPSGFGKPAR 411  
DB 385 TGEKPYKCTEGQKFSQSSALITHRRTYTGKPYQCGCGNF----- 427  
QY 412 GGVGKGFCSORODLKSPP-----SOVAHLRAHYLHTGKPYPCPTCGTRFR 461  
DB 428 ---SSSNLATHRRRLHLEKPYKGLCKGKSSQSSSLNTHQHTGKPYECLTGCSFS 484  
QY 462 HLQTLKSHVRIHTGKPYHPCDGLPHRHSQRLHLRQKIG 503  
DB 485 WSNLIKHQRTYTGKPYRCGCGKGFSGRQDLVHGHQTHNG 526

RESULT 8  
ZF93 MOUSE STANDARD, PRT, 645 AA.  
ID ZF93 MOUSE  
AC 06116;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Zinc finger protein 93 (Zfp-93).  
GN ZFP93.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.



DR PROSITE; P550157; ZINC\_FINGER\_C2H2\_2; 15.  
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
NM Nuclear protein; Repeat; Repressor.  
FT DOMAIN 4 75 KRAM.  
FT ZN\_FING 146 588  
FT ZN\_FING 146 168 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 174 156 C2H2-TYPE.  
FT ZN\_FING 202 224 C2H2-TYPE.  
FT ZN\_FING 230 252 C2H2-TYPE.  
FT ZN\_FING 258 280 C2H2-TYPE.  
FT ZN\_FING 286 308 C2H2-TYPE.  
FT ZN\_FING 314 336 C2H2-TYPE.  
FT ZN\_FING 342 364 C2H2-TYPE.  
FT ZN\_FING 370 392 C2H2-TYPE.  
FT ZN\_FING 426 448 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 454 476 C2H2-TYPE.  
FT ZN\_FING 482 504 C2H2-TYPE.  
FT ZN\_FING 510 532 C2H2-TYPE.  
FT ZN\_FING 538 560 C2H2-TYPE.  
FT ZN\_FING 566 588 C2H2-TYPE.  
FT CONFLICT 84 84 R -> Q (IN REF. 2; AA552689).  
FT CONFLICT 115 115 T -> I (IN REF. 2; AA58671).  
FT CONFLICT 177 177 T -> R (IN REF. 2; AA58671).  
FT CONFLICT 184 184 G -> R (IN REF. 2; AA58671).  
SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D62D43B CRC64;  
  
Query Match 15.2%; Score 424; DB 1; Length 595;  
Best Local Similarity 38.5%; Pred. No. 4,1e-15;  
Matches 90; Conservative 26; Mismatches 92; Indels 26; Gaps 5;  
  
QY 267 KCGAPASTPYLLTQADTSGSPSEPARPLPGSEFPSCQNCVAVAGCSSGL--DSLVPG 323  
DB 206 EGG-----KAFNMSSTLTGKRIHTGKPYKCEGAFNQSNNLTKHKIHTG 254  
QY 324 DEDKPYKCOLCRSSFFRYKGNLASHRTVHTGKPYKCSICGAPFNPANLKTSHRIHSGK 383  
DB 255 --EKPYKCEGCAFNRFSSTLTGKRIHTGKPYKCEGCAFNRSSSTLTGKRIHTGK 312  
QY 384 PYKCEGSAFVQVRSGPFGCKRARGVGQKGFCSNQRODLKSPSQVAHLPAHL 443  
DB 313 PYKCEGSAFVQVRSGPFGCKRARGVGQKGFCSNQRODLKSPSQVAHLPAHL 443  
QY 444 HTGKPYKCPCTGTRFRLIQLTKSHVRIHTGKPYKCHDPCGLHFRKSQLRLH 497  
DB 363 HTGKPYKCEGCAFNRFSSTLTGKRIHTGKPYKCEGCAFNRSSSTLTGKRIHTGK 416  
  
RESULT 10  
ID 2239 HUMAN STANDARD; PRT; 458 AA.  
AC 016600;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Zinc finger protein 239 (Zinc finger protein MOK-2) (HOK-2).  
GN ZNF239 OR MOK2 OR HOK2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OC NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96139011; PubMed=8587123;  
RT "Ernoul-Lange M., Arranz V., le Coniat M., Berger R., Kress M.;  
RT Human and mouse Kruppel-like (MOK2) orthologue genes encode two  
RT different zinc finger proteins.";  
RL J. Mol. Evol. 41:784-794(1995).  
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potentia).  
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X82126; CAA57638.1; -  
DR EMBL; X82125; CAA57637.1; -  
DR HSSP; P25490; IUBD.  
DR Genew; HGNC:13031; ZNF239.  
DR MIM; 601069; -  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 5.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR PRODOM; PD000003; Znf\_C2H2; 7.  
DR SMART; SM00355; Znf\_C2H2; 9.  
DR PROSITE; P500028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; P550157; ZINC\_FINGER\_C2H2\_2; 9.  
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
NM Nuclear protein; Repeat.  
FT DOMAIN 235 453 ZINC\_FINGERS.  
FT ZN\_FING 235 257 C2H2-TYPE.  
FT ZN\_FING 263 285 C2H2-TYPE.  
FT ZN\_FING 291 313 C2H2-TYPE.  
FT ZN\_FING 319 341 C2H2-TYPE.  
FT ZN\_FING 347 369 C2H2-TYPE.  
FT ZN\_FING 375 397 C2H2-TYPE.  
FT ZN\_FING 403 425 C2H2-TYPE.  
FT ZN\_FING 431 453 C2H2-TYPE.  
SQ SEQUENCE 458 AA; 51534 MW; 7749A0629A9D8920 CRC64;  
  
Query Match 15.1%; Score 421.5; DB 1; Length 458;  
Best Local Similarity 30.5%; Pred. No. 4,3e-15;  
Matches 107; Conservative 32; Mismatches 129; Indels 83; Gaps 10;  
  
QY 176 SCGQPPSPAPSPDPACN-----WKRYIVLNSQA-----SQAGLVGRSS 218  
DB 137 TQNG-QLKESLDIPDCKCKDHWGK--SQVSCGQGHTEKPCDINCKILNTSPD 193  
QY 219 GQPCQVRLPSGDEASSSSSSSSSEGGPIPGQSRLSPATVATVFKCAPASTPYLL 278  
DB 194 GHPYKIKITAEKQVGGSCGKNFSQSS-LL 223  
QY 279 TQADDTGSPSEPARPLPGSEFPSCQNC-EAVAGCSSGLDVLVGDDEDPKPYKCOLCRSS 337  
DB 224 LHQDHTKRP-----YKCEGCGKFTSSSLIHQAVHTDEKPYKCEGCGK 271  
QY 338 FRYKGNLASHRTVHTGKPYKCSICGAPFNPANLKTSHRIHSGKPYKCEGSAFVQ- 396  
DB 272 FTRSSSLIHAVHTGKPYKCDKCGKFGSQSKLHQRVHTGKPYKCEGSAFVQSR 331  
QY 397 ---VRSQPPSGFGCKRARGVGQKGFCSNQRODLKSPSQVAHLPAHLIHTGKPYK 452  
DB 332 SWLHIHQDHTG--EAPYKCEGCGK-----SQSSVLHIRCHTGTGKPYK 376  
QY 453 CPCTGTRRHLDLTKSHVRIHTGKPYKCHDPCGLHFRKSQLRLHQLQNG 503  
DB 377 CYECGKFGSQSSDLRIHVRHTGKPYKCGKFGSQSKLHQRVHTG 427  
  
RESULT 11  
ID 2145 HUMAN STANDARD; PRT; 673 AA.  
AC 005516;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Zinc finger protein PLZF (Promyelocytic leukemia zinc finger  
DE protein) (Zinc finger protein 145).  
GN ZNF145 OR PLZF.  
OS Homo sapiens (Human).  
OS -----

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart ventricle;  
 RX MEDLINE=93209216; PubMed=8384553;  
 RA Chen Z., Brand N.J., Chen A., Chen S.-J., Tong J.-H., Wang Z.-Y.,  
 RA Waxman S., Zelen A.;  
 RT "Fusion between a novel Kruppel-like zinc finger gene and the  
 RT retinotic acid receptor-alpha locus due to a variant t(11;17)  
 RT translocation associated with acute promyelocytic leukemia.";  
 RL EMO J. 12:1161-1167(1993).  
 RN [2]  
 RP SEQUENCE OF 424-455 FROM N.A.  
 RX MEDLINE=93253074; PubMed=8387545;  
 RA Chen S.-J., Zelen A., Tong J.-H., Yu H.-Q., Wang Z.-Y., Derre J.,  
 RA Berger R., Waxman S., Chen Z.;  
 RT "Rearrangements of the retinotic acid receptor alpha and promyelocytic  
 RT leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a  
 RT patient with acute promyelocytic leukemia.";  
 RL J. Clin. Invest. 91:2260-2267(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 7-122.  
 RX MEDLINE=20005701; PubMed=10537309;  
 RA Li X., Peng H., Schultz D.C., Lopez-Guisa J.M., Rauecher F.J. III,  
 RA Marmorestein R.;  
 RT "Structure-function studies of the BTB/POZ transcriptional repression  
 RT domain from the promyelocytic leukemia zinc finger oncoprotein.";  
 RL Cancer Res. 59:5275-5282(1999).  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. MAY PLAY A ROLE IN  
 CC MYELOID MATURATION AND IN THE DEVELOPMENT AND/OR MAINTENANCE OF  
 CC OTHER DIFFERENTIATED TISSUES.  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PLZF AND PLZPB (SHOWN HERE);  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WITHIN THE HEMATOPOIETIC SYSTEM, PLZF IS  
 CC EXPRESSED IN BONE MARROW, EARLY MYELOID CELL LINES AND PERIPHERAL  
 CC BLOOD MONONUCLEAR CELLS. ALSO EXPRESSED IN THE OVARY, AND AT LOWER  
 CC LEVELS, IN THE KIDNEY AND LUNG.  
 CC -1- INDUCTION: BY RETINOIC ACID.  
 CC -1- DISEASE: A FORM OF ACUTE PROMYLOCYTIC LEUKEMIA (APL) IS  
 CC CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;17)(Q32;Q21)  
 CC WHICH INVOLVES ZNF145 AND RETINOIC ACID RECEPTOR ALPHA (RAR).  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z19002; CA79489.1; -  
 DR EMBL: S60093; AAC60590.2; -  
 DR PDB: 1CG3; 09-AUG-00.  
 DR TRANSFAC: T02336; -  
 DR Genew; HGNC:12930; ZNF145.  
 DR MIM: 176797; -  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR000823; Znf\_C2H2.  
 DR Pfam: PF00096; zf-C2H2; 5.  
 DR Pfam: PF00651; BTB; 1.  
 DR PRINTS: PR00048; ZINCINGER.  
 DR ProDom: PD000003; Znf\_C2H2; 1.  
 DR SMART: SM00225; BTB; 1.  
 DR SMART: SM00355; Znf\_C2H2; 9.  
 DR PROSITE: PSS0097; BTB; 1.  
 DR PROSITE: PSS0028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE: PSS0157; ZINC\_FINGER\_C2H2\_2; 9.

KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat; Chromosomal translocation; Proto-oncogene;  
 KW Phosphorylation; Alternative splicing; 3D-structure.  
 FT DOMAIN 34 96  
 FT ZN\_FING 404 652  
 FT ZN\_FING 404 426  
 FT ZN\_FING 432 453  
 FT ZN\_FING 461 484  
 FT ZN\_FING 490 512  
 FT ZN\_FING 518 540  
 FT ZN\_FING 546 568  
 FT ZN\_FING 574 596  
 FT ZN\_FING 602 624  
 FT ZN\_FING 630 652  
 FT SITE 394 395  
 FT MOD\_RES 76 76  
 FT MOD\_RES 184 184  
 FT MOD\_RES 197 197  
 FT MOD\_RES 256 256  
 FT MOD\_RES 282 282  
 FT MOD\_RES 628 628  
 FT VARSELIC 255 377  
 FT SEQUENCE 673 AA; 74332 MW; 7CD7319E2A32109D CRC64;  
 Query Match 15.1%; Score 421.5; DB 1; Length 673;  
 Best Local Similarity 25.3%; Pred. No. 6.2e-15;  
 Matches 155; Conservative 63; Mismatches 236; Indels 159; Gaps 19;  
 20 HSSDVLGINELRLGLITDVTLLVGGPRLRAKAVLIACGFPYSIF-RGAGVGVDL 78  
 16 HPTGLCKANQRLAGTLCDDVIVWDSQEFHARVTLACTSKPFIILFRNSQHYTLDFL 75  
 79 SLPGCEARGFAPLDFMTSLRLSPATAPVLAATYLOMEHVQACHR---IQASY 135  
 76 S-----PFTQILLEVATYATLOAKABDLDDLAAILEILEYEBQCLKLTIOASD 129  
 136 E----- 136  
 130 DNDTEATMGAGAEEDBKAKAYLNIFISKSSSESGYASVAGSLPGPMVDQSPYST 189  
 137 -----PLGISLRPLENEPTPTAPPPGSPRRSEGHDPPTESRGS 178  
 190 SFGLSAMSPTKAAVDSMLTIGSLQGLTLPAGGEPEPTLAGGGHPPGVAEVTMMQVD 249  
 179 QGPSPASDPKACMKKKYKIVLNSQASQSLVGGSSQGPCQAR-----LPS 229  
 250 E---VPSQSPGAAL-----SSISGMDKVEERKSGGPTNSSVITSARELHY 297  
 230 GDEASS-----SSSSSSSSSEEGPIPGQSRISPTATVQFKGAPASTPYLLTS--- 280  
 298 GREBAEQVPPALAGQAPRTGPEHPAPPEHGLIYSLVLRPHKADAVLSMSSVTSGLH 357  
 281 --QAODTSSPSERARPLPGS---EFPSCQNCBAVACSSGLDLPVDEDEKPKYQCLC 334  
 358 VQPALAVSWDFSTYGLLPQGFITQRELPS-KLGEIYAVGWKS--ESRTIGEQ-----CSVC 409  
 335 RSPFRYKGLAHRVTHGKERHYSIGCARFNRPAHLKTHSRHIS-GEKPYKCTCSR 393  
 410 GVELPDNEAVEGHRKLHSGMKTYGCELCGKRLDSLRLMHLIASAGAKAVVCCQCAQ 469  
 394 F-----VQVRSQPPSG-----FOGKPA-----RGVQOKGFCSSQRODLKS 430  
 470 FSKEDALETHRQTHGTDMAVFCLLCGKRFQAQSLAQGHMEVNAVRSYICSECNRTF-- 527  
 431 PPSQVAHLRAVLIHTEKPYPCPTGTRFRHLQTLKSHVRIHTEKPYCDPCGLFRH 490  
 528 -FSHTA-LKRHLRSHTGHPYECFCGSCFRDESTLSKHKRIHTEKPYECNCGDKKPSL 585  
 491 KSQLHLHRLRQKIG 503  
 586 KQLETHYRVHTG 598

RESULT 12  
 ID HIC1 HUMAN STANDARD; PRT; 733 AA.  
 AC 014526;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN HIC1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID:9606;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE-96071505; PubMed-7585125;  
 RA Walee M.M., Biel M.A., el Dely W., Nelkin B.D., Isea J.-P.,  
 RA Cavenne W.K., Kuerbitz S.J., Baylin S.B.;  
 RT "p53 activates expression of HIC-1, a new candidate tumour suppressor  
 RT gene on 17p13.3.";  
 RL Nat. Med. 1:570-577 (1995).  
 [2]  
 RP ALTERNATIVE SPLICING, INTERACTION WITH CTBP, AND SUBCELLULAR LOCATION.  
 RX MEDLINE-21438979; PubMed-11554746;  
 RA Deltour S., Plute S., Guerdardel C., Leprieux D.;  
 RT "Characterization of HRC22, a human homologue of the putative tumor  
 RT suppressor gene HIC1.";  
 RL Biochem. Biophys. Res. Commun. 287:427-434 (2001).  
 CC -1- FUNCTION: Transcriptional repressor. May act as a tumor  
 CC suppressor. May be involved in development of head, face, limbs  
 CC and ventral body wall.  
 CC -1- SUBUNIT: Interacts with CtBP.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms: 1 (shown here) and 2;  
 CC are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed with highest levels  
 CC found in lung, colon, prostate, thymus, testis and ovary.  
 CC Expression is absent or decreased in many tumor cells.  
 CC -1- DISEASE: Defects in HIC1 may be a cause of the facial dysmorphism  
 CC and additional birth defects (except for lisencephaly) seen in  
 CC the contiguous gene disorder Miller-Dieker syndrome (MDS), like  
 CC defective development of the nose, jaws, extremities,  
 CC gastrointestinal tract, and kidney.  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 C2H2-TYPE ZINC FINGERS.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: I41919; AAD09201.1; -  
 CC HSSP: P08046; 1A1H.  
 CC Genew: HGNC:4909; HIC1.  
 CC MIM: 603825; -  
 CC DR InterPro: IPR000210; BTB\_POZ.  
 CC DR InterPro: IPR000822; ZnF\_C2H2.  
 CC DR Pfam: PF00651; BTB\_1.  
 CC DR Pfam: PF00956; ZF-C2H2; 5.  
 CC DR PRINTS: PR00048; ZINC\_FINGER.  
 CC DR ProDom: PD000003; ZnF\_C2H2; 1.  
 CC DR SMART: SM00325; BTB\_1.  
 CC DR SMART: SM00355; ZNF\_C2H2; 5.  
 CC DR PROSITE: PS00097; BTB\_1.  
 CC DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 CC DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 5.

KW Multigene family; Nuclear protein; Transcription regulation;  
 KW DNA-binding; Zinc-finger; Metal-binding; Repeat; Anti-oncogene;  
 KW Developmental protein; Alternative splicing.  
 FT DOMAIN 47 110  
 FT DOMAIN 160 167  
 FT DOMAIN 195 199  
 FT DOMAIN 439 613  
 FT ZN\_FING 439 459  
 FT ZN\_FING 509 529  
 FT ZN\_FING 537 557  
 FT ZN\_FING 565 585  
 FT ZN\_FING 593 613  
 FT SITE 241 245  
 FT VARSPLIC 1 19  
 SQ SEQUENCE 733 AA; 76467 MW; E8034428A4E490D3 CRC64;  
 MISSING (IN ISOFORM 2).  
 Query Match 15.1%; Score 420.5; DB 1; Length 733;  
 Best Local Similarity 23.9%; Pred. No. 7.6e-15;  
 Matches 151; Conservative 54; Mismatches 201; Indels 227; Gaps 20;  
 20 HSSDVLGNLNLRLRGLTDTLLVGGQPLRAHKAVLJACSGFPYSIFRGAGVGVDS 79  
 29 HSRQLQLQNNQRTGFLCDVITVQNALFRHKAVLAASAYLKSIV-----VADNLN 83  
 80 LFGGPEANG-FAPLIDFMYTSL-----RLSPATAP--AVLAATYIQMEHVQ 125  
 84 LDHDMVSPAVFRLVDFITGTGLADGAEAAAVAPGABPGLAVLAASVLTQIPDLVA 143  
 126 ACHRFITQ-----ASTYPEGISIRPLEAR-----PRTPTAP 157  
 144 LCKKRLKHKGYCHLRGGGGGGGYAPYGRPRGLRAATPVYQACYSRSPVGPAPPAEP 203  
 156 PGSPRS-----EGHPDP-----TESRSCS-----OGPPSPAPSD 188  
 204 PEGPAAVNTHCALYASGPGPAALCASERCSPLCGLDLSKSPPGSAAPERPLABRE 263  
 189 -----PKACNMKRYK-----YVLNSQASQAGSLVGRSSGQPCQARL 227  
 264 LPPRRDSPSAGPAAYKBPPLALPPLPFPQKLEMAPSPDPFRFGSGSPDEPRGRD 323  
 324 GSSLRYRMWKHEPGLGSGYDELGRGSPSECEBGRGDAAVSPGPGPLGLAPPRYGS 383  
 228 -PS-----GDEASSSSSSSSSSSSSEEG-----PIPGQSRLSPYATVQF 266  
 267 KCGAAR---STRYLLTSGAQTDSGSPS-----RARPLP-----GSFPECQNC- 307  
 384 LDGPGAGDGDYKSSSBETGSSBSPGGLLEGYPGPHLAYGEPZSPFGDNLVYCTPCG 443  
 308 -----EAVACSSGLDSTL 320  
 444 KQFPSSSEQLNAVEANVEEENLYGRAEAABAAGLPGPGGGGSKVAGAPGLGL 503  
 321 VEGDEDKPKYCOLCRSSFRYKGNLASHRTVTGKPYHCISICGAPFNNRPNLKTSHRS 380  
 504 L-----RPYRCASCCKSYKDPATLRQHKETHVLTLPYCTICGKFKTGTTRHRSRL 558  
 381 GEKPYKCEFCGSRFPQVNSQPPSGFGQKPARCGVQKGGFCSGQNDLSPSSQVAHLRA 440  
 559 GLKPPACACGCMRF-----TROYRLTE 580  
 441 HVLITGEEKPYPCPTGTRPRLIQTLLKSHVRIH 473  
 581 HMRITSGEKPYCQYCGGKFAQGRNLISHMKH 613  
 RESULT 13  
 ID HIC2 HUMAN STANDARD; PRT; 615 AA.  
 AC 096J3; Q9UPY3; Q9NSM9; Q9EKR3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypomethylated in cancer 2 protein (Hic-2) (Hic-3) (Hic1-related gene  
 on chromosome 22).  
 GN HIC2 OR HIC22 OR KIAA1020.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-65 FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.  
 RX MEDLINE=21438979; PubMed=11554746;  
 RA Deltour S., Pitre S., Guetardel C., Lepoint D.;  
 RT "Characterization of HIC22, a human homologue of the putative tumor  
 suppressor gene Hic1.";  
 RL Biochem. Biophys. Res. Commun. 287:427-434(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RA Alliel P.M., Goudou D., Bitoun M., Seddiq N., Rieger F., Perin J.-P.;  
 RT "Complete deduced structure of Hic-3, a novel human bcl/poz and Zp  
 factor of the Hic family.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Camp M., Smith L.J., Alencough R., Almeida J.P., Babbage A.K.,  
 RA Bagshaw C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrell M.D., Burton J., Carter C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockrill C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Levertha M.A., Lloyd C., Lloyd D.M.,  
 RA Maryn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA McClellan J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavlitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Senta H.K., Skuse C.D., Smalley S., Smith W.L.,  
 RA Soderlund C., Spurgeon L., Stewart C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki Y., Asakawa S., Kodoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuizawa S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chiswick S., Murray J., Miller N., Mink P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinde K., Kemp K., Lacroix P., Layman D., Ozerky P., Nelson T.,  
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurahashi H., Salter B.E., Edelmann L.,  
 RA McDevitt H.B., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,  
 RA Seroussi E., Franssen I., Tapia I., Brander C.E., O'Brien K.P.,  
 RA Wilkerson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tikhunov Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 RN [4]  
 RP SEQUENCE OF 61-615 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=9937452; PubMed=10470851;  
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:197-205(1999).  
 RN [5]  
 RP SEQUENCE OF 301-615 FROM N.A.  
 RC TISSUE=Testis;  
 RA Koehrer K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Transcriptional repressor.  
 CC -1- SUBUNIT: Interacts with CREB.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Highest levels in cerebellum.  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ13204; CAC70715.1; -;  
 DR EMBL: AF349035; AAK72951.1; -;  
 DR EMBL: AF000557; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: AB028943; BAA82972.1; -;  
 DR EMBL: AL162003; CAB82344.1; -;  
 DR Genew: HGNC:18595; HIC2.  
 DR Interpro: IPR000210; BTB\_POZ.  
 DR Interpro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00096; Zf\_C2H2; 5.  
 DR SMART: SM00355; Znf\_C2H2; 5.  
 DR PROSITE: PS50097; BTB; 1.  
 DR PROSITE: PS50028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Multigene family; Nuclear protein; Transcription regulation;  
 KW DNA-binding; Zinc-finger; Metal-binding; Repeat; Repressor;  
 KW Alternative splicing.  
 KM DOMAIN 46 109 BTB.  
 FT 442 615 ZINC\_FINGERS.  
 FT 442 469 C2H2-TYPE.  
 FT 505 532 C2H2-TYPE.  
 FT 533 560 C2H2-TYPE.  
 FT 561 588 C2H2-TYPE.  
 FT 589 615 C2H2-TYPE.  
 FT SITE 246 250 BINDING TO CTBP.  
 FT VARSPIC 18 MISSING (IN ISOFORM 2).  
 FT CONFLICT 49 51 ITM -> TIR (IN REF. 1).  
 FT CONFLICT 176 176 Q -> R (IN REF. 4).  
 SQ SEQUENCE 615 AA; 66155 MW; B03686C31B198C95 CRC64;  
 FO  
 Query Match 15.0%; Score 419; DB 1; Length 615;  
 Best Local Similarity 22.9%; Pred. No. 7,7e-15;  
 Matches 133; Conservative 64; Mismatches 205; Indels 246; Gaps 20;  
 QY 6 APEGALGVREPTRRSSDVLGNLNLRLRGILTDVTLVGGQPLRAKAVLIACGFFYS 65  
 DB 14 AARGDGMPEMELPSHSKQLQLQNGRKGLCDVIINWBNIPFAHGVLAASITPFS 73  
 QY 66 --IFRGAGGVGVVLSLPGEPARGFAPLDMYTSRLSLPATPA-----VLAATY 117  
 DB 74 LVLDNLNLINLTDVMS-----STVFQOILDFITYG--KLPSDDPAEPNFTLLTAASY 125  
 QY 118 LGMEHVQAQCHRFIQASVEPGLISRLPLEAEPPTPTAPPSPR-----RSEGH 168  
 DB 126 LQLEPULALCRKRLKAGKPFQ-----SGRAGSTGMGRPPRSQRLSTRASVIAQAYQGLV 179

```

QY 169 DPPTSRSCSGCPSPSPAPDPKACNMKKYTYLVNSQASQAGLVGERSGCPPOARIP 228
DB 180 DSRKGAHAFQELPOAKXSDD-----ELFLGGS---NQDSVQGLGRAVCP 220
QY 229 SGDEASSSSSSSSSSSEEG-----PI-----PGQSLSTFATAVQPKCA 270
DB 221 AGGEGAGLGGSSSTNGSGCGEOELDLKSKSPPLPAPTPGPHLTPDDAQLSDSOHS 280
QY 271 P--ASTPYLLTS-----280
DB 281 PPAASAPPVANASAYSELGTPDEPMDLGAEDNHLLEAPGCGPRKSLHSTRKMG 340
QY 281 QADOTSGSPSPAR-----PLGSEF-----PSCNCEA 309
DB 341 KKEPVAGSPFPERREARPKPCPEEGEGVDRVPNGILASGAGSPGYPEPPCKE--BE 399
QY 310 VAGCSGGLDLYPQDE-----DKPYKQLCGSSFRYKKNLAS 346
DB 400 ENGKDASEDSAGSGSGSGHASAHYMYROEGYETVSYDNLVYCIPCAAGPSSQLMA 459
QY 347 HRTVHTGE-----KPYHCSICGARFNRPAN 371
DB 460 HVEHTHEELFTKEGAVETGSGGAEBEADLSAPSAATTAERPRPKVCVCEKTYDPT 519
QY 372 LKTHSRHSGEKPYKCTCGSRFVYRSGPPSGFGKPARGVGKGFCSSQRODLKSP 431
DB 520 LKQHEKTHWLTTRPPFCNICGKMETQ-----RG-----546
QY 432 PQGVALLRAHVLIHNGEKPYPCPTGCTRRHLOTLKSHVRITNGEKPYHCDPGLHFRUK 491
DB 547 -TWTRHMRSHL---GLKPFACDECMRPTROYLRTEHMRVHSGEKPYECOLCGKFTQO 601
QY 492 SOLRLHLR 499
DB 602 RNLISHLR 609

RESULT 14
HIC1 CHICK STANDARD; PRT; 676 AA.
ID HIC1 CHICK Q90850; Q90851; Q90852;
AC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypomethylated in cancer 1 protein (Hic-1) (GammaFBP) (Fragment).
GN HIC1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Embryonic lens;
RX MEDLINE=94374565; PubMed=8088434;
RA Liu Q., Shalaby F., Puri M.C., Tang S., Breitman M.L.;
RT "Novel zinc finger proteins that interact with the mouse gamma
RT F-cysterealin promoter and are expressed in the sclerotic during early
RT somatogenesis.";
RT Dev. Biol. 165:165-177(1994).
CC -1- FUNCTION: Binds specifically to the gamma F-1-binding motif of the
CC gamma F-cysterealin promoter. May have a regulatory role in
CC sclerotic specification and/or differentiation. Isoform 2
CC functions as a transcriptional repressor in lens cells.
CC -1- SUBUNIT: Interacts with CtBP (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; 1/GammaFBP-A,
CC 2/GammaFBP-B and 3/GammaFBP-C (shown here); are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: Isoform 1 is highly expressed in kidney and
CC lung. Expression of isoform 2 is higher in the lens, retina and
CC stomach, and extremely low in heart, muscle, kidney and lung.
CC Isoform 3 is weakly expressed in heart, kidney and lens.

```

```

CC -1- DEVELOPMENTAL STAGE: In the embryo of stage 11, expressed
CC predominantly in the head mesenchyme surrounding the brain and in
CC the paraxial mesoderm. Highly expressed in pre-somitic mesoderm and
CC then over the entire epithelial somite. During somitic
CC differentiation, expression becomes restricted to the sclerotic.
CC In the developing lens, expression is most active at the beginning
CC of lens fiber cell differentiation.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79011; CAA55644.1; -
CC DR EMBL; X79050; CAA55652.1; -
CC DR EMBL; X79051; CAA55653.1; -
CC DR HSSP; P08046; 1A1H.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR000822; Znf_C2H2.
CC DR Pfam; PF00651; BTB; 1.
CC DR Pfam; PF00096; zfc_C2H2; 5.
CC DR PRINTS; PR00048; ZINCFINGER.
CC DR PRODOM; PD000003; Znf_C2H2; 1.
CC DR SMART; SM00225; BTB; 1.
CC DR PROSITE; PS00097; BTB; 1.
CC DR PROSITE; PS00355; Znf_C2H2; 5.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 5.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
CC DR Nucleic acid protein; Transcription regulation; DNA-binding; Zinc-finger;
CC Metal-binding; Repeat; Developmental protein; Repressor;
CC KW Alternative splicing.
CC FT NON TER 1
CC FT DOMAIN 1 63 126 BTB.
CC FT DOMAIN 2 420 585 ZINC FINGERS.
CC FT ZN FING 420 585 C2H2-TYPE.
CC FT ZN FING 420 585 C2H2-TYPE.
CC FT ZN FING 474 501 C2H2-TYPE.
CC FT ZN FING 502 529 C2H2-TYPE.
CC FT ZN FING 530 557 C2H2-TYPE.
CC FT ZN FING 558 585 C2H2-TYPE.
CC FT SITE 241 245 BINDING TO CTBP.
CC FT VARSPLIC 1 35 APGARPARSRKRGKSRERECGGAAGRARGA -> MR
CC FT VARSPLIC 1 35 VHRLEGLAEGSGRARGARGA (IN ISOFORM 1).
CC FT SEQUENCE 676 AA; 73758 MW; 3D3C7FD3302F32EC CRC64;

Query Match 15.0%; Score 419; DB 1; Length 676;
Best Local Similarity 23.7%; Pred. No. 8.4e-15;
Matches 146; Conservative 67; Mismatches 184; Indels 218; Gaps 21;

QY 9 GALGVREFTSRSSVGLNLMELRGLTDTVTLVGGQPLRAHRAVLIAGCFYSIFR 68
DB 34 GAMLEAMEVPSHSROLTLQTLNRTKGLCDYIVQNALFRANNTILAAASAVYKSLV- 92
QY 69 GRAGGVVDSLPGGPEARG-FAPLLDPMYTSRL-RLSPA--TAPAVLAATYLOMEHV 123
DB 93 ----VHDVLNLDHMSVPGIFRLDLFTYGRGCEGPGGSGVLAASVYQIDGL 148
QY 124 VOACHR-----PIQASYE--PLGISLRPL- 146
DB 149 VALCKKLLRSKGYCHLNGYAPYKLGRLATTPVIOACYSGTFRPVLDLPVEPAAPLN 208
QY 147 -----A 147
DB 209 TOGGEVLASAGCTPLPHGLCPERRHCSPPCGLDLSKSPGSAJLLPTDRLLPAEPR 268
QY 148 EPTTP-----TAPPGSPRSRSGP---DPPTSRSCSGPSP 184

```





```
Db 280 GAPELGLAVLAASYLOIPDLVALCKKLIKHKIKCHLRGGSGGGGAYAPYGRGLRA 339
Qy 148 E-----PPTPTAPPGSGP----- 161
Db 340 ATPVIQACSSPAGPPPPAAEPSPGPDAAVNTHCALYASGPPAASICAPERCSPLC 399
Qy 162 -----RSEEG 166
Db 400 GLDLCKSPGSSVPERPLSERELPPRPDPPGAGPAVYKEPSLALPPLPLPQKLEEA 459
Qy 167 HPDP-PTESRSCSQGPPSPAPDPKAC--NMKKYKIVLNSQASQASGVGERSS----- 218
Db 460 VPTDPFRSGSGSPGEPGPPGSGSLYRMKHE---?GLGSYGDELVRDRGSGPERL 515
Qy 219 ---GQPC-----PQARLP-----SGDEASSSSSSSSSEEGPIPGQGR 256
Db 516 EERGGDPAAAPGPPGLVPPPRYPGSLDGPGTGADGDDYKSSSEETGSSDDPSP-PCGH 574
Qy 257 LSPTAATVQFKCGAPASTPYLLTSAQD-----TSGSPS-----ERA 293
Db 575 LE-----GYPC--PHLAYGEPESEFGDNLVYCIPCGKGFPSEQLNAHVEAHVEEE 622
Qy 294 RPLPGESEFFSCNCEAVAGCCSGGLDVLVPGDEDK-----PYKQQLCRSSFRY 340
Db 623 EALYG---RAEAAEVAAG-AAAGLPFFGGGDKVTGAPGGLGELLRPYRCASCDKSYXD 677
Qy 341 KGNLASHRTVHTGEKPYHCSICGARFNRPANLKTSHRIHSGEKPYPKCTCGSRFYQVRSQ 400
Db 678 PATLQHEKTHWLTTPYCTICGKFTQGTMTTRHMRSHLGLKFPACDACGKRF----- 731
Qy 401 PPSGFQGKPARGGVGQKGGFCSSORODLKSPSQVAHLRAHVLHTGEKYPYCPCTGTRF 460
Db 732 -----TRQYRLTEHMRISHGEKPYECQVCGGKF 759
Qy 461 RHLQTLKSHVRIH 473
Db 760 AQQRNLISHMKH 772
```

Search completed: July 14, 2003, 18:16:07  
Job time : 8.38782 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 ; Search time 17.9594 Seconds  
(without alignments)  
3843.329 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789  
Sequence: 1 MGSPPAPGALGVREFTRH.....RQKGMATNTKVHYHILGSP 518

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgcdata/geneeq/geneeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2789	100.0	518	22	AAE11887
2	1138	40.8	707	23	ABBS7289
3	1122.5	40.2	706	21	AA178793
4	1122.5	40.2	706	22	AA296440
5	1095.5	39.3	706	16	AA68743
6	1095.5	39.3	706	21	AA78792
7	582	20.9	810	21	AA73351
8	576.5	20.7	803	19	AAW81756
9	574.5	20.6	803	22	AAW39272
10	574.5	20.6	804	22	AAW41058

11	483	17.3	610	23	AAW47790
12	482.5	17.3	611	22	AAW40035
13	482	17.3	610	22	ABBS0159
14	482	17.3	610	22	AAW93761
15	482	17.3	610	23	AAW47779
16	480	17.2	603	22	AAW41821
17	469.5	16.8	711	23	AAO21779
18	457.5	16.4	641	22	AAW84369
19	457.5	16.4	641	23	ABG61930
20	438.5	15.7	620	22	AAU30585
21	437	15.7	645	22	ABG28285
22	436.5	15.7	725	22	AAW95116
23	431.5	15.5	535	23	ABP41884
24	430	15.4	841	22	ABG16167
25	428.5	15.4	678	22	AAW25801
26	428	15.3	516	22	AAW61332
27	426.5	15.3	127	22	AAW95679
28	426.5	15.3	402	22	AAW93941
29	425	15.2	488	22	AAW69703
30	425	15.2	488	22	AAW95422
31	423	15.2	488	22	AAW93950
32	423	15.2	488	22	AAW93418
33	421.5	15.1	471	22	AAW93961
34	419.5	15.0	345	22	AAW64584
35	419	15.0	516	22	AAW95738
36	419	15.0	632	22	ABG19011
37	418	15.0	1050	22	ABG09685
38	417.5	15.0	569	22	AAW95128
39	417.5	15.0	569	22	AAW95733
40	417	15.0	406	22	AAU16023
41	415.5	14.9	480	23	ABBS7271
42	415.5	14.9	803	22	AAW95278
43	415	14.9	809	22	AAW38689
44	413.5	14.8	702	23	AAW04333
45	413.5	14.8	799	21	AAW21003

## ALIGNMENTS

RESULT 1  
AAE11887  
ID AAE11887 standard; Protein; 518 AA.  
AC AAE11887;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Angiogenesis associated human BAZF (hBAZF) protein.  
XX  
KW Angiogenesis associated protein; AAP; cytosolic; cardiac; gene therapy;  
KW ophthalmological; vulnary; myocardial infarction; muscular degeneration;  
KW diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;  
KW rheumatoid arthritis; psoriasis; drug screening; tumor; transplantation;  
KW cancer; therapeutic; diagnostic; human; BAZF protein.  
XX  
OS Homo sapiens.  
XX  
PN WO200170808-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 22-MAR-2001; 2001WO-US09609.  
XX  
PR 22-MAR-2000; 2000US-191134P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX (GENTH) GENENTECH INC.  
XX  
PI Rastelli LK, Gerritsen M;  
XX WPI, 2001-602775/68.  
XX N-PSDB; AAD19116.  
DR

Murine myoneurine.  
Human polypeptide  
Human transcrip  
Human protein sequ  
Human myoneurine.  
Human polypeptide  
Lung-specific amn  
Novel acid sequenc  
Prostate cancer-as  
Novel human secret  
Novel human diagno  
Human protein sequ  
Human ovarian anti  
Novel human diagno  
Human protein sequ  
Human protein sequ  
Human reproductive  
Human polypeptide,  
Human zinc finger  
Human protein sequ  
Human polypeptide,  
Human protein sequ  
Human polypeptide,  
Human zinc finger  
Human protein sequ  
Human zinc finger  
Human novel diagno  
Novel human diagno  
Human protein sequ  
Human polypeptide  
Human zinc finger  
Human nucleic acid

XX Novel angiogenesis associated polypeptides and polynucleotides encoding  
PT the polypeptides, useful for modulating angiogenesis and for treating  
PT tumors and cancers -  
XX  
PS Claim 1, Page 12-13, 159pp; English.  
XX  
CC The invention relates to angiogenesis associated proteins (AAP) and their  
CC corresponding cDNA molecules, which are useful for modulating  
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting  
CC wound healing, for example after organ transplantation, and in the  
CC treatment of tumours, myocardial infarction, cancers, diabetic  
CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
CC AAP proteins and DNA's are useful in potential prophylactic and  
CC therapeutic applications implicated in a variety of disorders including  
CC those related to angiogenesis, and also in diagnostic applications.  
CC AAP cDNA is also useful in gene therapy. The invention also relates to  
CC a method for screening a tissue sample for tumourigenic potential. AAP  
CC proteins are used to screen drugs or compounds that modulate AAP activity  
CC or expression as well as treating disorders characterised by insufficient  
CC or excessive production of AAP or production of AAP forms that have  
CC decreased or aberrant activity compared to the wild type protein, or  
CC modulate biological function that involve AAP. The present sequence  
CC is human BAZF (hBAZF) protein which is an angiogenesis associated  
CC protein (AAP) of the invention. BAZF is a Bcl-6 (LAZ3) homolog, a  
CC transcription repressor that controls germinal center formation and the  
CC T-cell dependent immune response. Human BAZF plays a role in cell  
CC proliferation.  
XX  
SQ Sequence 518 AA;

Query Match 100.0%; Score 2789; DB 22; Length 518;  
Best Local Similarity 100.0%; Pred. No. 6, 8e-189;  
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60  
DB 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60  
QY 61 GFYYSIFRGRAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYIOM 120  
DB 61 GFYYSIFRGRAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYIOM 120  
QY 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQG 180  
DB 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQG 180  
QY 181 PPSBPASDPYACMKWKYKTYVLNSOASQAGSLVGRSSGQPCQARLPSCDEASSSSSS 240  
DB 181 PPSBPASDPYACMKWKYKTYVLNSOASQAGSLVGRSSGQPCQARLPSCDEASSSSSS 240  
QY 241 SSSSSSEGPFRGOSRLSPATATVOPFCGAPASTPYLLTSQADOTSSPERARPIPGSE 300  
DB 241 SSSSSSEGPFRGOSRLSPATATVOPFCGAPASTPYLLTSQADOTSSPERARPIPGSE 300  
QY 301 FFSQONCEAVAGSSGGLDSLVPGEDDKPYKQLCRSFRYKGMASHRTVHTGEKPYHGS 360  
DB 301 FFSQONCEAVAGSSGGLDSLVPGEDDKPYKQLCRSFRYKGMASHRTVHTGEKPYHGS 360  
QY 361 ICGARFRPANKLTHSRHSGEKPYKCEGTSFRVQYRSQPSGFGQKPARGVGQKGF 420  
DB 361 ICGARFRPANKLTHSRHSGEKPYKCEGTSFRVQYRSQPSGFGQKPARGVGQKGF 420  
QY 421 CSSGORDLSPBSOVANLRAHVLIHNGEKYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480  
DB 421 CSSGORDLSPBSOVANLRAHVLIHNGEKYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480  
QY 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
DB 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
QY 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
DB 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518

RESULT 2

ABB57289.  
ID ABB57289 standard; Protein; 707 AA.

AC ABB57289;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:814.

KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

PN MO200188188-A2.

PD 22-NOV-2001.

PF 18-MAY-2001; 2001WO-0P04192.

PR 18-MAY-2000; 2000JP-0145977.

PA (UNTI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

DR N-PSDB; AB199745.

PT Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring  
PT expression levels of particular genes defined in the specification or  
PT by determining the expression profile of a gene group comprising these  
PT genes -

PS Claim 2, Page 2004-2007; 2630pp; English.

XX  
CC The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (AB199202 to AB199912, encoding  
CC the protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The  
CC expression levels or expression profiles produced by these genes are  
CC used as an indicator when screening for ischaemic condition improving  
CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
XX which are used in the exemplification of the present invention.

SQ Sequence 707 AA;

Query Match 40.8%; Score 1138; DB 23; Length 707;  
Best Local Similarity 36.4%; Pred. No. 4e-72;  
Matches 269; Conservative 51; Mismatches 149; Indels 270; Gaps 14;

QY 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60  
DB 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60  
QY 61 GFYYSIFRGRAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYIOM 120  
DB 61 GFYYSIFRGRAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYIOM 120  
QY 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQG 180  
DB 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQG 180  
QY 181 PPSBPASDPYACMKWKYKTYVLNSOASQAGSLVGRSSGQPCQARLPSCDEASSSSSS 240  
DB 181 PPSBPASDPYACMKWKYKTYVLNSOASQAGSLVGRSSGQPCQARLPSCDEASSSSSS 240  
QY 241 SSSSSSEGPFRGOSRLSPATATVOPFCGAPASTPYLLTSQADOTSSPERARPIPGSE 300  
DB 241 SSSSSSEGPFRGOSRLSPATATVOPFCGAPASTPYLLTSQADOTSSPERARPIPGSE 300  
QY 301 FFSQONCEAVAGSSGGLDSLVPGEDDKPYKQLCRSFRYKGMASHRTVHTGEKPYHGS 360  
DB 301 FFSQONCEAVAGSSGGLDSLVPGEDDKPYKQLCRSFRYKGMASHRTVHTGEKPYHGS 360  
QY 361 ICGARFRPANKLTHSRHSGEKPYKCEGTSFRVQYRSQPSGFGQKPARGVGQKGF 420  
DB 361 ICGARFRPANKLTHSRHSGEKPYKCEGTSFRVQYRSQPSGFGQKPARGVGQKGF 420  
QY 421 CSSGORDLSPBSOVANLRAHVLIHNGEKYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480  
DB 421 CSSGORDLSPBSOVANLRAHVLIHNGEKYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480  
QY 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
DB 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
QY 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518  
DB 481 CDECGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518

```

QY 152 ----- 151
DB 235 SAROVNEYSRPAMEVSPSLCHSNINYSPEKAVPEARSIDHVSPEGPAPVANSANAPY 294
QY 152 ----- 176
DB 295 PCDKASKEERPSSEDEIALHPEPPNAPLNRKGLVSPQSPKSDQCPNPTSCSKNA 354
QY 177 ----- 225
DB 355 CILQASGSPPAKSPDPKACNMKKYKFIYVLSNLQNAKSGSEQALGRLSPRAYAPPA 414
QY 226 RLPSPGDEAS-SSSSSSSSSSSEGPPIGPQSR-----SPTATVQF----- 266
DB 415 COPPEPAPLIDQSPFKLSASGEDSTIP-QASRLNLIYNRSLGSPRSSSESHSPLYMHP 473
QY 267 -KCGAPASTPYLLTQAODTSGSPSERAPLPSESE-----FSCNCEAVAGCSSG 316
DB 474 PKCTSCGSGSPQHTKCHTACPTPEEMGETOSEYSDSSCENGTFPCNCCDRFSEAS 533
QY 317 LD-SLVPGDEDKPYKCOLCRSFRYKGNLASHRTVHTGEKPYHCISGARFNRPAVLKTH 375
DB 534 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHKTHTVTEKPYRCNICGAQFNRPAVLKTH 593
QY 376 SRHSGEKPKCETCGSRFVQVRSPSGFQGRPARGVGQKGFCSQRODLKSPSQV 435
DB 594 TRHSGEKPKCETCGARFV-----QV 615
QY 436 AHLRAHVLHTGEKPYPCPTCGRFRHLQTLKSHVRIHTGEKPYHCDPCGLHFRKSOAR 495
DB 616 AHLRAHVLHTGEKPYPCICGRFRHLQTLKSHLRIHTGEKPYHCEKCNLHFRKSOAR 675
QY 496 LHLRQHGATNTKVHYHI 514
DB 676 LHLRQHGALITNTKVQYRV 694

RESULT 3
AAV78793
ID AAV78793 standard; protein; 706 AA.
XX
AC AAV78793;
DT 19-MAY-2000 (first entry)
XX
DE Human BCL-6 protein sequence.
XX
KW Bcl-6; human; B-cell lymphoma; regulator; non-Hodgkin's lymphoma;
KW diffuse type B-cell lymphoma.
XX
OS Homo sapiens.
XX
PN WO200000185-A1.
XX
PD 06-JAN-2000.
XX
PF 30-JUN-1999; 99WO-US14703.
XX
PR 30-JUN-1998; 98US-0107058.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Dalla-Favera R, Niu H;
XX
DR WPI; 2000-160631/14.
XX
PT Novel methods for regulating BCL-6 levels in cells used to treat humans
PT with lymphoma
XX
PS Example 2; Fig 10; 159pp; English.
XX
CC This sequence represents the human bcl-6 protein sequence. The invention
CC relates to a vertebrate bcl-6 locus which is the breakpoint cluster

```

region in B-cell lymphomas, and containing a bcl-6 gene encoding a BCL-6 polypeptide. Administration of a molecule which induces phosphorylation of BCL-6 and thereby induces BCL-6 degradation, can be used as a method of regulating BCL-6 in cells. The method of the invention can be used to regulate, and especially to decrease BCL-6 levels in cells. The method may also be used to screen putative therapeutic agents for treatment of non-Hodgkin's lymphoma, by contacting cells from lymphoma and normal cells with the agent, and after a period of time comparing the amount of bcl-6 nucleic acid in each sample, a difference indicating the effectiveness of the agent. The bcl-6 gene is a source of probes and primers, which are used to diagnose diffuse-type B cell lymphoma and B cell lymphoma in a subject. Anti-BCL-6 antibodies may also be used for this purpose. The methods are useful for treating non-Hodgkin's lymphoma.

XX Sequence 706 AA;

Query Match 40.2%; Score 1122.5; DB 21; Length 706;  
Best Local Similarity 36.3%; Pred. No. Se-71;

Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;

```

QY 1 MGSAPAPGALGVBEFTRHSSDVGNINELRLGILDTVLVGGQPLRAKAVLIACS 60
DB 1 MASPA--DSCT-----QFTRHASDVLIANRLRSRIILDVIVVSRBQRAKTYLMACS 54
QY 61 GPFYSIFGRAGVGVVLSLPGEPARGAPLLDMYTSRLSLSPATAPVLAATYLLQM 120
DB 55 GLFYSIFPDQLKCNLSVINLDEINPEGCILDMYTSRLNLRGHTMAVATMYLLQM 114
QY 121 EHVVAQCHRFIOAS----- 134
DB 115 EHVVDTCRFIVASEAENVSAIKPREBEFLNSRMLPODIMAYRGRIVENNLPRLSABG 174
QY 135 -----YEPGIS----- 141
DB 175 CESRAPASLVSGLSTPPASYSMTSHLPVSLISFDEBRDYRMVAPPPERALLPCDS 234
QY 142 -----LAPL----- 145
DB 235 ARPVGEXRPTLEVPNVCHSNINYSPKETIPBEARSDMHWASVAGLKAAPASANAPYF 294
QY 146 -----EAEPPT-----PPTAP-----PGSPRSSEGHFDPPTSRSS----- 176
DB 295 PCDKASKEERPSSEDEIALHPEPPNAPLNRKGLVSPQSPKSDQCPNPTSCSKNAK 354
QY 177 -----CGQPPSPASPDPKACNMKKYKFIYVLSNLQNAKSGSEQALGRLSPRAYAPPA 414
DB 355 ILQASGSPPAKSPDPKACNMKKYKFIYVLSNLQNAKSGSEQALGRLSPRAYAPPA 414
QY 224 QARLPSPGDEASSSSSSSSSSEGPPIGPQSR-----SPTATVQF----- 266
DB 415 QPPM-EPENLIDQSPFKLSASGEDSTIP-QASRLNLIYNRSLGSPRSSSESHSPLYMHP 472
QY 267 ---KCG-APASTPYLLTQ---AODTSGSPSERAPLPSESEFSCNCEAVAGCSSG 316
DB 473 PKCTSCGSGSPQHAEMCLHTAGPTFAEEMGETOSEYSDSSCENGTFPCNCCDRFSEAS 532
QY 317 LD-SLVPGDEDKPYKCOLCRSFRYKGNLASHRTVHTGEKPYHCISGARFNRPAVLKTH 375
DB 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHKTHTVTEKPYRCNICGAQFNRPAVLKTH 592
QY 376 SRHSGEKPKCETCGSRFVQVRSPSGFQGRPARGVGQKGFCSQRODLKSPSQV 435
DB 593 TRHSGEKPKCETCGARFV-----QV 614
QY 436 AHLRAHVLHTGEKPYPCPTCGRFRHLQTLKSHVRIHTGEKPYHCDPCGLHFRKSOAR 495
DB 615 AHLRAHVLHTGEKPYPCICGRFRHLQTLKSHLRIHTGEKPYHCEKCNLHFRKSOAR 674
QY 496 LHLRQHGATNTKVHYHI 514
DB 675 LHLRQHGALITNTKVQYRV 693

```

RESULT 4  
 AAB29640  
 ID AAB29640 standard; Protein; 706 AA.  
 XX  
 AC AAB29640;  
 XX  
 DT 23-FEB-2001 (first entry)  
 XX  
 DE Human bcl-6 transcriptional repressor.  
 XX  
 KM Human; bcl-6; transcriptional repressor; germinal centre formation;  
 KM Th-2 mediated antibody affinity maturation; apoptosis regulator;  
 KM chromosome 3q27; lymphoma; acute lymphoblastic leukaemia;  
 KM post-transplant lymphoproliferative disorder; expression inhibition;  
 KM anticancer therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6140125-A.  
 XX  
 PD 31-OCT-2000.  
 XX  
 PF 15-OCT-1999; 99US-0418640.  
 XX  
 PR 15-OCT-1999; 99US-0418640.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI Taylor JK, Cowseert LM;  
 XX  
 DR WPI; 2001-048959/06.  
 DR N-PSDB; AAC81137.  
 XX  
 PT Anticancer compounds which specifically hybridize with and inhibit human  
 PT bcl-6 expression, useful for treating bcl-6 related disorders, and  
 PT preventing or delaying inflammation or tumor formation  
 XX  
 PS Disclosure; Column 47-52; 42pp; English.  
 XX  
 CC This sequence represents human bcl-6. Bcl-6 (also known as B-cell  
 CC CLL/lymphoma 6, zinc finger protein 51 and LZ23) is a sequence-  
 CC specific DNA-binding transcriptional repressor. The bcl-6 gene is  
 CC expressed in germinal centre B- and T-cells and is required for  
 CC germinal centre formation and Th-2 mediated antibody affinity maturation.  
 CC Bcl-6 may also play a role in the regulation of apoptosis. The bcl-6 gene  
 CC is located on chromosome 3q27, a region which undergoes a high frequency  
 CC of translocation events. Such chromosomal translocations can result in  
 CC aberrant forms of bcl-6, which are strongly implicated in the  
 CC pathogenesis of several types of lymphoma, and have also been reported  
 CC in acute lymphoblastic leukaemia and post-transplant lymphoproliferative  
 CC disorders. The invention relates to antisense oligonucleotides targeted  
 CC to the human bcl-6 gene, which inhibit its expression. A series of  
 CC oligonucleotides (AAC114-C8123) were designed to target different  
 CC regions of the human bcl-6 mRNA, and were analysed for their effect on  
 CC bcl-6 mRNA levels by quantitative real-time PCR. The oligonucleotides of  
 CC the invention are useful for diagnosis, prevention and treatment of  
 CC conditions associated with aberrant forms of bcl-6, such as lymphomas,  
 CC acute lymphoblastic leukaemia and post-transplant lymphoproliferative  
 CC disorders.  
 CC  
 XX  
 SQ Sequence 706 AA;  
 XX  
 Query Match 40.24; Score 1122.5; DB 22; Length 706;  
 Best Local Similarity 36.34; Pred. No. 5e-71;  
 Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;  
 QY 1 MGSPAAEGALGYRETRHSSDVLGNLRLRGILTDVTLVVGQPLRAHRAVLAACS 60  
 DB 1 MASPA--DSCI-----GTRHSDVLTNLRSLRDLITDVIVVSEQPRAHKTVLMACS 54  
 QY 61 GPFYISIFRGAGVGVVLSLPGGEARGFAPLDDFWYTSRLSPATAPVLAAYTLQW 120  
 DB 55 GLFYISIFTDQLKCNLSVINLDPINPEGFCILDDFWYTSRLNREGNIMAVMATAYLQW 114

QY 121 EHVQACHRFIOAS----- 134  
 DB 115 EHVQDCKRFIKASEHMSAIKPREBEFLNSRMLMPQIMAYRGREVNNLPLRSARG 174  
 QY 135 -----YEPGLGIS----- 141  
 DB 175 CESRAFAPSLVGLSTPPASYSMSHLPVSSILPDSDEPRDVMVAPNPPYERALPCDS 234  
 QY 142 -----LRPL----- 145  
 DB 235 ARPVEEYSRPLEVSPNVCNSNIYSPKTIPEBARSDMHVAVAGLKAASARNAPYF 294  
 QY 146 -----EAEPPT-----PPTAP-----PPGSPRSEGHDPPTESRS----- 176  
 DB 295 PCDKASKEEERSSEDEIALHFEPPNAPLNKRGVLPQSPQSDQPNPTEACSSKXAC 354  
 QY 177 ---CSGGPPSPASPDPKACNMKKYKVIYNS-----QASQAGSLVGRSSGQPCP 223  
 DB 355 ILQASGSPPAKSPDPKACNMKKYKVIYNSLNQNAKPGGPGQAEIGRLSPRAYTAPAC 414  
 QY 224 QARLPBGDEASSSSSSSSSEBGPDPGOSRL-----SPTATVQF----- 266  
 DB 415 QPPM-EPENLDIQSPKUSASGEDSTTP-QASRLNNIVNRSMTGSPRSSSEHSPLVMP 472  
 QY 267 ---KCG--APASTPYLITSQ-----AODTSGSPERAPLPSEBFFSCQNCBAVAGCSG 316  
 DB 473 PKCTSGSGSPQHAEMCLHTAGPTFAEBEGERTOSVSDSCENGAFPCNECDRSESEAS 532  
 QY 317 LD-SLVPGDEDPYKQOLCRSSFRYKGNLASHRTYHTGKPYHCISICGARFRRPANKTH 375  
 DB 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHRTYHTGKPYRCNICGQAFRRPANKTH 592  
 QY 376 SIHSGEKYKCTGSGRPVQVRSQPPSGFGKPARCGVGQSGRQDLKSPSQV 435  
 DB 593 TIHSGEKYKCTGSGRVR-----QV 614  
 QY 436 AILRAVLIHTGEKYPCTGTRFRHLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQUR 495  
 DB 615 AILRAVLIHTGEKYPCTGTRFRHLQTLKSHRIHTGEKPYHCEKNLHFRHKSQUR 674  
 QY 496 LHLRQHGATNTKQVYHI 514  
 DB 675 LHLRQHGATNTKQVYRV 693  
 RESULT 5  
 AAR68743  
 ID AAR68743 standard; Protein; 706 AA.  
 AC AAR68743;  
 XX  
 DT 19-JUL-1995 (first entry)  
 XX  
 DE BCL-6 zinc finger protein.  
 XX  
 KM bcl-6 locus; non-Hodgkin lymphoma; B-cell lymphoma; B-lymphocyte;  
 KM diagnostic; therapeutic; chromosome-3q27; translocation;  
 KM proto-oncogene; diffuse large cell lymphoma; DLCL; zinc finger.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 520..541  
 FT /label= Zinc-finger  
 FT Region 548..568  
 FT /label= Zinc-finger  
 FT Region 576..596  
 FT /label= Zinc-finger  
 FT Region 604..624  
 FT /label= Zinc-finger  
 FT Region 632..652  
 FT /label= Zinc-finger

```

FT Region 660..681
FT /label= Zinc-finger
XX MO9429343-A.
XX
XX 22-DEC-1994.
XX
XX 09-JUN-1994; 94WO-US06669.
XX
XX 09-JUN-1993; 93US-0074967.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX (SLOK ) SLOAN KETTERING INST CANCER.
XX
XX Chaganti RK, Dalla-Favera R, Chaganti RS;
XX
XX WPI, 1995-036403/05.
XX N-PSDB; AAQ68743.
XX
XX Nucleic acid from genetic locus bcl-6 - used to develop probe.
XX for diagnosis and therapy of B-cell lymphoma and non-Hodgkin's
XX lymphoma
XX
XX Disclosure; Page 90-94; 129pp; English.
XX
XX DNA was extd. from tumor tissue of 2 cases of IGH-producing diffuse-
XX type B-cell NHL carrying the t(3;14) (q27;q32) translocation. DNA
XX analysis showed that the breakpoints on 3q27 were located within 3
XX kb of the same genomic locus, which was designated bcl-6. The
XX human bcl-6 locus was identified by screening a phage cDNA library
XX constructed from Bjab B-cell lymphoma mRNA. The zinc finger
XX protein encoded by bcl-6 is given in AAR68743. ECL-6 is a
XX proto-oncogene specifically involved in the pathogenesis of diffuse
XX large cell lymphoma.
XX
XX Sequence 706 AA:
XX
XX Query Match 39.3%; Score 1095.5; DB 16; Length 706;
XX Best Local Similarity 35.7%; Pred. No. 4e-69;
XX Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;
XX
XX 1 MGSPAAEGALGVREFTRRSSDVLGNLRLGILTDVTLVGGQPLRAHKAVALIACS 60
XX 1 MASPA--DECI---QFTRHARDVLTNLRSLRDITDVIVVSRSEGRFAKTVLMAMR 54
XX
XX 61 GFFYSIFRKGAVGVVLSLPGEPARGAPLIDFMYTRSLRLSPATAVLAATYLM 120
XX 55 GLFYSIFRQDKCNLSVINLDEPEINDEGCILDFMYTRSLRLREGNINAVATMYLM 114
XX
XX 121 EHVVOACHRFIONS-----YEP LGIS----- 134
XX 115 EHVVDTCRKFIOKSEAMVSAIKPPEEFNLNMLPMODIMAYRGHEVENLPLRSAG 174
XX
XX 135 -----YEP LGIS----- 141
XX 175 CESRAPAPSLVGLSTPPASYSWYSHLPVSLFSDPEFRDVMVPAVNPFRKRALPCDS 234
XX
XX 142 -----LRL----- 145
XX 235 ABPVGESVPTLEVPNVCHSNIVSPKTEPREARDNMYSVABGLKPAASARNAVYF 294
XX 146 -----BAEPPT-----PPTAP-----PGSPRSEBHPDPTES---RSC 177
XX 295 PCDKAKSEKERRSSBRIALHPEPPVAPLNRKGLVPSQSPYSDCCPNSPTEKCSKXAC 354
XX
XX 178 ----SGPPSPAPDPDKACNMKKYKIVINS-----QASQAGSLVGERSSGQPCP 223
XX 355 ILQSGSPPAKSPPTDEKACSMKKYKIVINSLNQNAKPGQBEQALGRSLSPRAYTAPAC 414
XX
XX 224 QARLPBGCDASSSSSSSSSSSEGRIPRQGRNL-----STATAVQF----- 266
XX 415 QPPM-EPENLDLQSPYTKLSASGSDSTIP-QASRLNINIVNR3MTGSPRSSSESHSLYMP 472

```

```

XX 267 ----KCG--APASTPYLLTSQ-----ADPTSGSPSERARPLGSEPFSSCONCEAVAGCSSG 316
XX 473 PKCTSCGSGSPQAHMCMHTAGPTFAEMGEGTQSEYSSSCENGAFFNCEDCDRSBEAS 532
XX
XX 317 LD-SLVPDEDEKPYKQCLCRSSFRYKGNLASHRYHTGEKPYHSGICGARPRRPNLKT 375
XX 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHRYHTGEKPYRCNICAOFRPRANLKT 592
XX
XX 376 SRHSGEKPYKCECTGSRFVGVRSQPPSGFGKPARGVGQKGGFCSSQRDLKSPFSGV 435
XX 593 TRHSGEKPYKCEITGAFV-----QV 614
XX
XX 436 AMLRAHVLIHTGEKPYKCEPTGTRFRHLQTLKSHVRIHTGEKPYHCDPGLHFRKSQLR 495
XX 615 AMLRAHVLIHTGEKPYKCEICGTRFRHLQTLKSHVRIHTGEKPYHCKCNHFRKSQLR 674
XX
XX 496 LHLRKGAGATNTKYNHII 514
XX 675 LHLRKGAGATNTKQYRV 693
XX
XX RESULT 6
XX AA78792
XX ID AA78792 standard; Protein; 706 AA.
XX
XX AA78792;
XX
XX 19-MAY-2000 (first entry)
XX
XX Human BCL-6 protein sequence.
XX
XX Bcl-6; human; B-cell lymphoma; regulator; non-Hodgkin's lymphoma;
XX diffuse type B-cell lymphoma.
XX
XX Homo sapiens.
XX
XX WO200000185-A1.
XX
XX 06-JAN-2000.
XX
XX 30-JUN-1999; 99WO-US14703.
XX
XX 30-JUN-1998; 98US-0107058.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Dalla-Favera R, Niu H;
XX
XX WPI; 2000-160631/14.
XX N-PSDB; AA290110.
XX
XX Novel methods for regulating BCL-6 levels in cells used to treat humans
XX with lymphoma -
XX
XX Examples; Fig 9; 159pp; English.
XX
XX This sequence represents the human bcl-6 protein sequence. The invention
XX relates to a vertebrate bcl-6 locus which is the breakpoint cluster
XX region in B-cell lymphomas, and contains a bcl-6 gene encoding a BCL-6
XX polypeptide. Administration of a molecule which induces phosphorylation
XX of BCL-6 and thereby induces BCL-6 degradation, can be used as a method
XX of regulating BCL-6 in cells. The methods of the invention can be used to
XX regulate, and especially to decrease BCL-6 levels in cells. The method
XX may also be used to screen putative therapeutic agents for treatment of
XX non-Hodgkin's lymphoma, by contacting cells from lymphoma and normal
XX cells with the agent, and after a period of time comparing the amount of
XX bcl-6 nucleic acid in each sample, a difference indicating the
XX effectiveness of the agent. The bcl-6 gene is a source of probes and
XX primers, which are used to diagnose diffuse-type B cell lymphoma and B
XX cell lymphoma in a subject. Anti-BCL-6 antibodies may also be used for
XX this purpose. The methods are useful for treating non-Hodgkin's lymphoma.

```

SQ Sequence 706 AA;  
 Query Match 39.3%; Score 1095.5; DB 21; Length 706;  
 Best Local Similarity 35.7%; Pred. No. 4e-69;  
 Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

QY 1 MGSPAAEGALGYREFTTHSSDVLGNLNLRLGILTDVTLVGGQPLRAHKAVALIACS 60  
 1 MASPA--DSCI-----QTRHARDVLTNLNLRSDILTDVIVVSREQFRAHKTIVLMAWR 54  
 DB 61 GFPEYSLFRGAGGVVDVLSLPGCEAEGFAPLDFMTSTLSRLSPATPAVLAATYTLQM 120  
 55 GLFYSLFTDQKCNLSVINLDPINPEGFCILLDFMTSTSLNREGNIMAVMATAYLQM 114  
 QY 121 EHYVQACHRFIOAS----- 134  
 115 EHYVDTCRKFTKASAEAMVSAIKPRREFLNSRLMMDIMAYRGREVENNLPLASAPG 174  
 DB 135 -----YEPGIS----- 141  
 175 CESRAFAPLSYGLSTPPASYSWYSHLPVSSLLFSDDEFDVMVPANPPKERALPCDS 234  
 QY 142 -----LRPL----- 145  
 DB 235 ARPVGESRPTLEVPNCHSNISPKETIPPEARSDMKVSAEGLKPAAPARAPAYF 294  
 QY 146 -----EAEPT-----PPTAP-----PPGSRSEGHDPPTES-----RSC 177  
 DB 295 PCOKASKEERPSSEEDIALHFEPPNAPLNRKGLVBPQSFQKDCQPNSTELCSKMKC 354  
 QY 178 ----SOGPSAPSPDKACNWKYKTYVLNS-----QASQAGSLVGERSSGQPCP 223  
 DB 355 ILGSGSPAPKSPDTPKACSMKKYKFTVLSLNQNAKGGPEQAEIGRLSPRAYTAPAC 414  
 QY 224 QALPBGDEASSSSSSSSSEEGPIPGQSR-----SPRAIVQF----- 266  
 DB 415 QPBM-EPENDLQSPYTLASGSDSTIP-QASRLNINIVNSMTGSPSSSESHSPMYMP 472  
 QY 267 ----KCG--APASTPYLLTSQ-----AQTGSPSERARPLPGSEFSGQCEAVAGCSCG 316  
 DB 473 PKTSCSGOSPQIAEMKLTHTAGPTFAEMGETSEVSDSCENGAFPCNECDRFEEMS 532  
 QY 317 LD-SLVPGDEKPYKCOLCRSSFRYKGNLASHRTVTGKPYKCSICGAFNRPANLKT 375  
 DB 533 LKQHTLQTHSDKPYKCDRCQASFRYKGNLASHRTVTGKPYKCNICGAFNRPANLKT 592  
 QY 376 SRHSEKPYKCTCGSRFYQVRSQPPSGFGKPARAGVQKPGCSQKDLKSPSQY 435  
 DB 593 TRHSGEKPYKCTCGARFV-----OV 614  
 QY 436 AHLRAVLITHTGKPYPCPTCGTRFRHLQTLKSHVRLHTGKPYKCDPCGLAFRHKSQ 495  
 DB 615 AHLRAVLITHTGKPYPCPTCGTRFRHLQTLKSHVRLHTGKPYKCEKCNLHRHKSQ 674  
 QY 496 LHLRQKGAATNTKVAHYI 514  
 DB 675 LHLRQKGAATNTKVAHYI 693  
 RESULT 7  
 AAY73351 standard; Protein; 810 AA.  
 XX AAY73351;  
 AC AAY73351;  
 DT 24-FEB-2000 (first entry)  
 DE HTRM clone 1484257 protein sequence.  
 XX HTRM: human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
 KM arteriosclerosis; cirrhosis; cancer; leukemia; diabetes mellitus;  
 KM Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
 KM trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.

XX OS Homo sapiens.  
 XX PN MO9957144-A2.  
 XX PD 11-NOV-1999.  
 XX PF 04-MAY-1999; 99WO-US09935.  
 XX PR 05-MAY-1998; 98US-0084254.  
 XX PR 07-AUG-1998; 98US-0095827.  
 XX PR 02-OCT-1998; 98US-0102745.  
 XX PA (INCY-) INCYTE PHARM INC.  
 PI Hillman JL, Bandman O, Lai P, Yue H, Reddy R, Tang YT;  
 PI Gerstein EH, Patterson C, Baughn MR, Azimzai Y, Lu DM;  
 DR MPI; 2000-052941/04.  
 DR N-PSDB; AA52436.  
 XX PT New peptides useful for diagnosis, prevention and treatment of cancer  
 PT and immune disorders.  
 PS Claim 1; Page 112-114; 193pp; English.  
 XX AAY73325-Y73389 are human transcriptional regulator molecule (HTRM)  
 CC protein sequences. The HTRM protein and nucleotide sequences are useful  
 CC for preventing or treating disorders associated with decreased expression  
 CC or activity of HTRM which include cell proliferative disorders such as  
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
 CC leukemia; immune disorders such as AIDS; Addison's disease, diabetes  
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
 CC of the HTRM polypeptides are useful for treating or preventing disorders  
 CC associated with increased expression or activity of HTRMs. HTRM  
 CC polypeptides, their immunogenic fragments or oligopeptides are useful  
 CC for screening libraries of compounds in drug screening techniques.  
 CC Polynucleotides encoding HTRM are useful for blocking the transcription  
 CC of mRNA and regulating gene function by modulating the activity of  
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or  
 CC treat disorder associated with decreased HTRM expression. Antibodies  
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
 CC for diagnosing disorders associated with the expression of HTRM,  
 CC particularly in assays that detect the expression of HTRM. Nucleotide  
 CC sequences encoding HTRM may be useful to generate hybridization probes  
 CC useful in mapping the naturally occurring genomic sequence and to detect  
 CC differences in gene sequences among normal, carrier and affected  
 CC individuals. Using diagnostic assays, cancer can be detected prior to  
 CC the appearance of clinical symptoms and thereby progression of cancer can  
 CC be prevented by aggressive treatment or preventive measures.  
 XX

SQ Sequence 810 AA;  
 Query Match 20.9%; Score 582; DB 21; Length 810;  
 Best Local Similarity 27.3%; Pred. No. 8.8e-33;  
 Matches 175; Conservative 69; Mismatches 182; Indels 216; Gaps 21;

QY 16 EFTRHSSDVLGNLNLRLGILTDVTLVGGQPLRAHKAVALIACGFFSTIRGAG-VG 74  
 DB 2 DPHQSHVLEQDNLNOORQGLGCTCFVVDGVAHFAHKAVALIACSEYFMLEVDKDVH 61  
 QY 75 VDVLSLPGCEAEGFAPLDFMTSTLSRLSPATPAVLAATYTLMEHYVQACHRFIOAS 134  
 DB 62 LDI-----SNAAGGVLEFMTYKLSLSPENVDDVLAATVFLQMODIITACH----- 109  
 QY 135 YEPGISLRLPEAEPTP-----TAPPGSPRSE-----GH 167  
 DB 110 -----ALKSL-AEATSPGCVNAEALAQKVCVPSGDKAKAEKVATSTLSRLQGR 162  
 QY 168 PPPPESR-----SCSG-----PPSPASPPKACNWKYKTYVLN 203  
 DB 163 STPIGSRDLKEERGQASASGAEQTEKADAPREPPEVLEKDPPTS-----G 211





```

XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000MO-US34263.
PF
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX MPI; 2001-442253/47.
DR N-PSDB; AAI58428.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2417; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilization of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 803 AA;
SQ

```

```

Query Match 20.6%; Score 574.5; DB 22; Length 803;
Best Local Similarity 29.4%; Pred. No. 3e-32;
Matches 180; Conservative 64; Mismatches 205; Indels 163; Gaps 20;

```

```

QY 16 EFTRHSSDVLGNINELRLGILTDVTLVGGQPLRAHKVLYLSCGFYSIFGRAG-VG 74
DB 2 DFGHSGHVLQOLNQRGLCDCTFVDGVHFKAKVLAACSEIFKMLFVDQKOVH 61
QY 75 VDVLSPGGEARGFALDPMYTSRLSPATAPVLAATYLAQHEHVAQCHRIQAS 134
DB 62 LDI-----SNAAGLGQVLEFMTYAKLSIPENVDDVLAATFLQMDITACH----- 109
QY 135 YELVGLISRLPEAPPP-----TAPPGSPRSEGHDPPTESRSCGQPPSPASP 188
DB 110 -----ALKSL-AEPTISPGNAEALATEGCDRAKEEKVATSLRLQAGRSTPGR- 161
QY 189 PKACNMKKYIYVLSNOSAGSLVBERSGCPQARLP---SGDEASSSSSSSSSSSS 245
DB 162 --SRDLKEERGGAGQASASAEOTEKADAPREPPVELKDPPTSGMAAAEAELSESSE 219
QY 246 BEGPI-EGPOSRLSPATATVQFCGA-PAST-----PYLLTSQA-----QDTSGSPS--- 290
DB 220 QMEVEVEPARKGEEQKEQEEGEGAPAEVKEEGSGLNGEAPENENESAGTDSQGE 279
QY 291 --ERAPLP-----GSEF-----FSCQNC 307

```

```

DB 280 LGSEARGLSGTGDRTESKAYSVIHKCEDCKEFTHTGNFKRIHHTGKPPSCREC 339
QY 308 EAV----AGCSSGLDVLVGEDDKPYKCOLCRSPRY----- 340
DB 340 SKAFSDPPACKAKHEKTHSP--LKPYGCEBCKSLRLISLNLHKKRHSGEARYCEDCG 396
QY 341 -----KGNLASHRTVHTGKPYHCISICG-----AREN 367
DB 397 KLFITSGNLKRLQLVHSGEKPYQCDYCGRSFSDPTSKRMHLETHDTEHKPCPHCKPN 456
QY 368 RANLKTSHRHSNGEKPYKCFCTGSRFPVYRSQPPSSGQKARAGVQKGFSCSROD 427
DB 457 QVGNLKAHLKIHADGPKRCRCGKQFT-----TSGNLKRLRIHSGEKPYVCIHCRO 510
QY 428 LKSPSOVAHLRAVLIHTGKPYPCPTGTRFRLIOTLKSIVRIHTGKPYHCDPCGLH 487
DB 511 PADPGA---LQRAVRIHTGKPYCCGWCGRKFTQASSLIHVRQHTGKPYVCRCGRK 566
QY 488 FHHKSQRLHLR 499
DB 567 FVSSQLANHIR 578

```

```

RESULT 10
AAM41058
ID AAM41058 standard; Protein; 804 AA.
XX
XX AAM41058;
AC
XX 22-OCT-2001 (first entry)
DE
XX Human polypeptide SEQ ID NO 5989.
XX
XX Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000MO-US34263.
PF
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX MPI; 2001-442253/47.
DR N-PSDB; AAI60214.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5989; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

```

CC Immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilization of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

CC Sequence 804 AA;

Query Match 20.6%; Score 574.5; DB 22; Length 804;  
 Best Local Similarity 29.4%; Pred. No. 3e-32;  
 Matches 180; Conservative 64; Mismatches 205; Indels 163; Gaps 20;

QY 16 EFTRRSSDVLGNLNLRLGLITDVTLLVGGQPLRAHKAVLAACSGFYISIRGRAG-VG 74  
 DB 3 DEPOHSQVLEBQUNQKQGLCDCTFVVDGVHFKAHKAVLAACSEYFMLEFVDQDVH 62  
 QY 75 VDVLSLPGPBARGFAPLDFMYTSRLSPATAPAVLAATYQMEHVQAQHPFIQAS 134  
 DB 63 LDI-----SNAGAGQVLEFMYTATLSPEVNDVLAATFLQODITICH----- 110  
 QY 135 YEPGLISLAPLEAPPTP-----TAPPGSPRSEGHDPPTESRSCSQGPSPASPD 188  
 DB 111 -----ALKSL-AEPATSPCGNAEALATCGDRABEKAVATSLRLQAGRTPIGP- 162  
 QY 189 PRACMKKTKYTVLNSQASQAGSLVGERSSGQPCPARLP-----SGDEASSSSSSSSSS 245  
 DB 163 --SRDKERGGQAGSAAAGAEQTEKADAPREPPPELKPDPPTSGMAAEAALESSE 220  
 QY 246 EEGPI-PGPQSRISPTATVQPKCA-PAST-----PYLLTSCA-----QDTSGPS- 290  
 DB 221 GEMVEVPARKGEBOKEQEBEGAGPAVEEGSQLHNGEAPENENEBESAGTDSQGE 280  
 QY 291 --ERARPLP-----GSEF-----PSCQNC 307  
 DB 281 LGSSEARGLSGYGDRTESKAYGVHKGEDCKEFTHTGTFKRIHITGKEPFCREC 340  
 QY 308 EAV-----AGCSGSLDVLVDEDEKPKCQLCRSSFPY----- 340  
 DB 341 SKAFSDPAKCAHEKTHSP--LKPYGCEBCKSKYRLISLNLKRRHSGEARYCEDCG 397  
 QY 341 -----KGNLASHRTVTGGEKPYHCSICG-----ARPN 367  
 DB 398 KLPFTSGNLRHQLVHSGEKPYQCDYCGSFDPSTKMELETHTDIXHKHCPCMKKN 457  
 QY 368 RPANLKTSHRISGEEKPYKCEGSRFVOVRSQPSGFCQKPARGVGQKGFCSQROD 427  
 DB 458 QVGNLKAHLKIHADPDLKCREGKQFT-----TSGNLKRLHRLISGEEKPYVCICHQ 511  
 QY 428 LKSPSQVAVLRAHVLHITGEEKPYPCPTGTRPRHQLKSHVRLITGGEKPYHCDPCGLH 487  
 DB 512 FADPGA---LQNHVAVHITGEEKPYCQVCWCAKAPTQASSLIHVRQHTGEEKPYCERCGR 567  
 QY 488 FRHKSQRLRLR 499  
 DB 568 FVQSSQLAHNR 579  
 RESULT 11  
 AAMA7790  
 ID AAMA7790 standard; Protein; 610 AA.  
 XX AAMA7790;  
 AC  
 XX  
 DT 01-MAR-2002 (first entry)

XX Murine myoneurine.  
 DE Murine; myoneurine; cytostatic; neurological; muscular; gene therapy;  
 XX cancer; neuromuscular disease; dystrophy; polymyositis; rhabdomyosarcoma.  
 OS Mus sp.  
 PN FR2808535-A1.  
 PD 09-NOV-2001.  
 XX 05-MAY-2000; 2000FR-0005792.  
 XX 05-MAY-2000; 2000FR-0005792.  
 PR 05-MAY-2000; 2000FR-0005792.  
 PA (ALLI/) ALLIEL P M.  
 PI Seddiql N, Clifuentes DC, Rieger F, Perin JP;  
 XX WPI; 2002-019508/03.

PT New nucleic acid encoding human myoneurin, useful for diagnosis,  
 PT prognosis and treatment of cancer and neurological diseases, comprises  
 a polynucleotide encoding the human myoneurin protein -

Claim 13; Page 61; 64pp; French.

CC The present sequence is the protein sequence for murine myoneurin. The  
 CC myoneurin coding sequence and its fragments are useful as primers and  
 CC probes, for diagnosis, prognosis or assessing risk of human diseases,  
 CC particularly cancers and neuromuscular diseases (e.g. dystrophy,  
 CC polymyositis or rhabdomyosarcoma). The coding sequence and its fragments  
 CC are also useful in gene therapy for these diseases.

CC Sequence 610 AA;

Query Match 17.3%; Score 483; DB 23; Length 610;  
 Best Local Similarity 27.2%; Pred. No. 6.2e-26;  
 Matches 152; Conservative 56; Mismatches 215; Indels 136; Gaps 14;

QY 16 EFTRRSSDVLGNLNLRLGLITDVTLLVGGQPLRAHKAVLAACSGFYISIRGRAG--GV 73  
 DB 2 QVSHHCLEHLERLNKQRBAGFLCDCTVIGFQFYAHNVLASFSEYGAIVRSTSENNV 61  
 QY 74 GYDVLSLGGPBARGFAPLDFMYTSRLSPATAPAVLAATYQMEHVQAQ----- 127  
 DB 62 FLDSQV---KADFQKLEFITYGTNLDSWVKEIHQADYLVKEVVYKCKIKMED 117  
 QY 128 HRFIQASYEPLGISLRLPEAPPTPTAPPPCS-----PRRSGHPPTESRSC 177  
 DB 118 FAFI-----ASPSREISITGNTILNQOALCLTATRTNNREKSEVSDS 162  
 QY 178 SQGPSPASPPEKAMKMKYKTVLNSQASQAGSLVGERSSGQPCPARLPSPGDEASSSS 237  
 DB 163 VQANPKPALTKKSSQSKKKKAFSSQPGQ-----SKAVQPSDVLASAVELFLD 214  
 QY 238 SSSSSSSSEEGIPPPQS-----RLSTATV-----QPCG----- 269  
 DB 235 TSKLSSPVEQIITQNDSELELTSVENTPEPTODIVTVTVKRRRKSQSHCALKEISM 274  
 QY 270 --APASTPYLLTSCAOD-----TSGSPSRARPL-----PGSEFFSCON 306  
 DB 275 SNIAVSKSPYLEMNGEFLDQRFSAKAMCMTCGAVFSEASSLRHMRHIVKRYVCHL 334  
 QY 307 C-EAVAGCSGSLDVLVDEDEKPKYKQLCRSSFRYKGNLASHRTVHTG-EKPYHCSICGA 364  
 DB 335 CGKAFQCNQLKTHVTRHTGSRPYKCELDGFAKQCLVPHSRMHSEBKPYKCDVCL 394  
 QY 365 RPNRANLKTSHRISGEEKPYKCEGSRFVOVRSQPSGFCQKPARGVGQKGFCSQ 424  
 DB 395 QPATSSNLKIHARKHSGEEKPYVCDRCGRF----- 424

QY 425 RODLKSPSPQVAHLRAHVLHITGKPYPCPTGTRFRHLQTLKSHVRIHTGKPYHCDPC 484  
Db 425 -----AQAStLTyVRHRTGKEKPYCDTGKAFVSSSLTTHSKRTGKEKPYICGIC 476  
QY 485 GLHFRHKSQRLRLRQKHG 503  
Db 477 GKSFSSGELNKHFRSHTG 495

RESULT 12  
AAM40035  
ID AAM40035 standard; Protein: 611 AA.  
AC AAM40035;  
XX  
XX  
XX 22-OCT-2001 (first entry).  
DE Human polypeptide SEQ ID NO 3180.  
XX  
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
PD 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000MO-US34263.  
PF  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI59191.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
PS  
XX Example 4; SEQ ID NO 3180; 10078bp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemia and  
XX C.N.S disorders.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification.

XX SQ Sequence 611 AA;  
Query Match 17.3%; Score 482.5; DB 22; Length 611;  
Best Local Similarity 26.9%; Pred. No. 6, 8e-26;  
Matches 151; Conservative 71; Mismatches 238; Indels 101; Gaps 17;  
QY 16 EFTRRSHSDVLGNLNLRLGLITDVTLLVGGQPLRAHKAVALIACGFFYSIFRGA--GV 73  
Db 2 QYSHHCEHLERLNLNQRBAAGFLCDCTIYIGFQFAHRNVLASFSEYGAIVRSTSENNV 61  
QY 74 GVDVLSLPGGPPARGAPFLDPMYTSRLRLSPATPAVLAATYIQMEHYVOACH----- 128  
Db 62 FLDSQV---KADGPKLLEFIYTGTLNLSMWVKEIHQADYLVKEVYVTKXIKMED 117  
QY 129 -----RFIOASGEPGLGILRP-----LEAPPPPTAPRPG 159  
Db 118 FAFIANPSTETISITGNTIELNQOTCLTLRDYNNREKSEVSTDILQANPKGALAKSS 177  
QY 160 -SPRSSEGHDPPTESRSCSQP-----PSPASPPKACNWK--KY 197  
Db 178 QTKKKKKAFNSPKTGQNTKQVQPSDILENASVELFLDANKLPTPVVEQVAQINDNSELEL 237  
QY 198 KTIYVNSQASQ--AGSLVGERSSGQPCPOARLP--GDEASSSSSSSSSEEGPIPG 252  
Db 238 TSVVENTFPAQDVIHVTIVTKRKRGKSPICALKSHSMNSIASVKSPEYEAENSCEE--LDQ 295  
QY 253 POSRLSPATAVQVFCGAPASTFPVLLTSGAODTSGSPSERAR---LPGSFFSCQNC 308  
Db 296 RYSKAKPMCNT---CGKYFSASSLRRMRHKG---VKPYCHLCGKAFTCNOULK 346  
QY 309 AVAGCSSLGLSLVPDEDEKPYKQCLRSSFRYKGNLASHRTVHTG-EKPYHSGICGAREN 367  
Db 347 THVRTHTG-----EKPYKCELCDKGFAQKQVLFRSHHGEKPYKCDVNLQGA 397  
QY 368 PRANLKTSHRITSGEKPYKCECTCGSRFQVNSQPSGSGQKARCGVQKGFSCSSROD 427  
Db 398 TSSNLKIHARKHSGEKPYKCDRCQGRFAQA---STLYHVRALLKEKPYVCTCGKA 452  
QY 428 LKSPSPQVAHLRAHVLHITGKPYPCPTGTRFRHLQTLKSHVRIHTGKPYHCDPCGLH 487  
Db 453 FAVSSSLITHSK---HTGKPYICGICGKSFSSGELNKHFRSHTGERTPCICGNS 508  
QY 488 FRHKSQRLRLRQKHGAATNT 508  
Db 509 YTDINLKKHKTKVSGADKT 529

RESULT 13  
ABB50159  
ID ABB50159 standard; Protein: 610 AA.  
XX  
XX ABB50159;  
XX  
XX 05-FEB-2002 (first entry)  
XX  
XX Human transcription factor TRFX-10.  
DE  
XX  
XX Human; transcription factor; TRFX; cell proliferative disease;  
KW autoimmune disease; inflammation; neurological disease;  
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;  
KW neuroprotective; antiinflammatory; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX WO200172777-A2.  
XX  
XX 04-OCT-2001.  
XX  
XX 13-MAR-2001; 2001WO-US08117.  
XX  
XX 13-MAR-2000; 2000US-0188986.  
XX

PA (INCY-) INCYTS GENOMICS INC.  
 XX Hillman JL, Baughn MR, Yue H, Lai P, Lu DAM, Patterson C;  
 PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;  
 PI Reddy R;  
 XX WPI; 2001-570896/64.  
 DR N-PSDB; ABA82983.  
 XX  
 PT Novel transcription factor polypeptides, used to treat diseases  
 PT associated with altered activity and expression of TRFX, and to screen  
 PT for agents capable of modulating its activity -  
 XX  
 PS Claim 1; Pages 150-151; 327pp; English.  
 CC  
 CC The present sequence is the protein sequence for a human transcription  
 CC factor. The transcription factor and its coding sequence are useful in  
 CC the diagnosis, treatment and prevention of diseases associated with  
 CC altered expression of the transcription factor e.g. cell proliferative,  
 CC autoimmune/inflammatory, neurological and developmental disorders. A  
 CC number of specific disorders/diseases are given in the specification,  
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.  
 XX  
 SQ Sequence 610 AA;  
 Query Match 17.3%; Score 482; DB 22; Length 610;  
 Best Local Similarity 26.8%; Pred. No. 7.3e-26;  
 Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;  
 QY 16 EFTRRSSDVLGNINELRLGILTDVTLVGGQPLRAHKAVALIACSGFFYSIFRGA--GV 73  
 DB 2 QVSHHEHLERLNKREAGFLCDCTIVGEFQKARHNLASFSEYFGALYSTENNV 61  
 QY 74 GVDVLSLPGSPPEARFAPLLDFMYTSRLSPATAVAVALAATYLGMEHYQACH----- 128  
 DB 62 PLDQSGV---KADGQKLEPFTYGTLDNSNVKEIHQADYLVVEEYVTKIKMED 117  
 QY 129 -----RPIQASYEPLGISLRP-----LEASPPTPTAPPG 159  
 DB 118 PAFIANPSSTEISSITGNIELNQOTLTLTRDYNNEKEVSTDLIQANPKQALAKKS 177  
 QY 160 -SPRSEGHDPPTESRSCSGP-----PSPASDPRAQCNWK--KY 197  
 DB 178 OTKKKKKAFNSPKTGQNKTVQYPSDILENASVELPLDANKLPPPVVEQVAQINDSELET 237  
 QY 198 KTIYVNSQASQ--AGSLVGBRSSGQPCQARLPS---GDEASSSSSSSSSEEPPIG 252  
 DB 238 TSVENTTTPPADIVHTVTVKRGKQPCALKEHSMNSNAYKSYEVAENSEE--LQO 295  
 QY 253 POSRLSPATATVQFKCAGAPASTPYLLTSQADTSGSPSEARP---LPGSEFFSCNCE 308  
 DB 296 RYSKAPMCNT---CGKVPSEASSLRMRIRIKG-----VKYVCHLGKAFQCNQJK 346  
 QY 309 AVAGSGSLDLSVGEDKPKYCOLCRSSFRYGNILASHRTVHTG- EKRYHCSI GARFN 367  
 DB 347 THVRTHTG-----EKRYKCELDCKGFAQKQCLVFHSRMHNGEKKPKCDVCMQLQFA 397  
 QY 368 RPAHLTHSRHSGEKKPKYCEGTSRFPVQVRSQPPSGFOCKRARGVGQKGGCSSOROD 427  
 DB 398 TSNLKIHKHKSGERPVCDRCGR----- 424  
 QY 428 LKSPPEQVLAHLRAVLIHTGEKPYPCPTGSTRFHLQTLKSHVRIHTGEKPYHDCGLH 487  
 DB 425 -----AQASTLTHTVHARHNGEKPYVCDTCKKAFVSSSLTTHSRKHTGEKPYICGLCGKS 479  
 QY 488 FRHKSQLRLHLRQKIG 503

DB 480 FISSEGLNKHRSHTG 495  
 RESULT 14  
 ID AAB93761  
 ID AAB93761 Standard; Protein; 610 AA.  
 XX  
 XX AAB93761;  
 AC  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:13450.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 XX  
 PD EPI074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-0116126.  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 13450; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the combination  
 CC of oligonucleotide comprises at least 15 nucleotides and the combination  
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 610 AA;  
 Query Match 17.3%; Score 482; DB 22; Length 610;  
 Best Local Similarity 26.8%; Pred. No. 7.3e-26;  
 Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;  
 QY 16 EFTRRSSDVLGNINELRLGILTDVTLVGGQPLRAHKAVALIACSGFFYSIFRGA--GV 73

```

Db      2 QVSHHCHLELRNKKQREAGFLCDCTIVIGEFQFAHRNVLASFSEYFGAIYRSTSENNV 61
      74 GVDVLSLPGGPEARGFAPLIDPMYTSRLRLSPATAPVLAATYLOMEHVQACH----- 128
      62 FLDSQV-----KADGFOKLLEFIYTGTLNDSMNWKEIHQADYLKVEEVVTKCKIKMED 117
      129 -----RFIOASYEPLGISLRP-----LEAPPTPTAPPG 159
      118 PAFIANPSTSEISITGTELNQOTCLTLRDYNNRKSSEVSTDLIQANPKGALAKSS 177
      160 -SPRSEGHDPPTESRSCSQGP-----PSPAPDPKACNMK---KY 197
      178 QTKKKKAFNSPKTGQNKTVQYPSDILENASVELFLDANKLPTPVVEQVAQINDSELEL 237
      198 KYIVLNSQASQ--AGSLVGERSSGQPCQARLPS---GDEASSSSSSSSSEEGPIPG 252
      238 TSVVENTFPAQDVIHTVTKRKKGKQOPNCALKEHSMNSIASVKSPLYEANSGBE--LDQ 295
      253 POSRLSPATATVQFCGAPASTPYLLTTSQAODTSGSPSERARP---LPSEFFSCQNC 308
      296 RYSKAKPMCNT---CGKVFSEASSLRHRIRHKG-----VKPYVCHLGGKAFQCNQLK 346
      309 AVAGCGSLDSLVPGEDDKPYKCOLCRSFRYKGNLASHRTVHTG-EKPYHCSICGARN 367
      347 THVRTHTG-----EKPYKCELCDGKFAKCOLVPHSRMHGEEKPKYCDVNLQGA 397
      368 RPAHLKTHSRINSGEKPKYKCECTGSRFVQVRSQPPSGFGKPARGGVGQKGFCSGORQ 427
      398 TSSNLKIHARKSGEKPKYVCDRCGQRF-----424
      428 LKSPSQVAHLRAHVLHTGEEKPYKCECTGSRFRHLQTLKSHVRINHTGKPYHCDPCGLH 487
      425 -----AQASTLYTHVRHRTGEEKPYVCDTCGKAFVASSSLITHSRKHTEKPYICGICGS 479
      488 FRHKSQRLHLRQKHG 503
      480 FISGELNKHFRSHTG 495

RESULT 15
ID AAM47779 standard; Protein; 610 AA.
AC AAM47779;
DT 01-MAR-2002 (first entry)
DE Human myoneurine.
KW Human; myoneurine; cytosolic; neurological; muscular; gene therapy;
KW cancer; neuromuscular disease; dystrophy; polymyositis; rhabdomyosarcoma;
KW BTB/POZ domain; zinc finger domain.
OS Homo sapiens.
FH Key 1.111 Location/Qualifiers
FT /note= "BTB/POZ domain. This sequence is also given
FT in AAM47780"
FT 304..527
FT /note= "zinc finger domain. This sequence is also given
FT in AAM47781"
PN FR2808535-A1.
PD 09-NOV-2001.
PF 05-MAY-2000; 2000FR-0005792.
PR 05-MAY-2000; 2000FR-0005792.
PA (ALLI/) ALLTEL P M.

```

```

XX Seddiq N, Cluentees DC, Rieger F, Perin JP;
PI MPI: 2002-019508/03.
XX DR N-PSDB; ABA04825.
XX DR N-PSDB; ABA04825.
XX PT New nucleic acid encoding human myoneurin, useful for diagnosis,
XX PT prognosis and treatment of cancer and neurological diseases, comprises
XX PT a polynucleotide encoding the human myoneurin protein -
XX PS claim 18; Fig 2; 64pp; French.
XX CC The present sequence is the protein sequence for human myoneurin. The
XX CC myoneurin coding sequence and its fragments are useful as primers and
XX CC probes, for diagnosis, prognosis or assessing risk of human diseases,
XX CC particularly cancers and neuromuscular diseases (e.g. dystrophy,
XX CC polymyositis or rhabdomyosarcoma). The coding sequence and its fragments
XX CC are also useful in gene therapy for these diseases.
XX SQ Sequence 610 AA;

Query Match 17.3%; Score 482; DB 23; Length 610;
Best Local Similarity 26.8%; Pred. NO. 7.3e-26;
Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;

16 EFTRHSSDVLGNLNEIRLGLTDTVLVGGQPLRAHKAVALIACGFFYSIRGRA--GV 73.
2 QVSHHCHLELRNKKQREAGFLCDCTIVIGEFQFAHRNVLASFSEYFGAIYRSTSENNV 61
74 GVDVLSLPGGPEARGFAPLIDPMYTSRLRLSPATAPVLAATYLOMEHVQACH----- 128
62 FLDSQV-----KADGFOKLLEFIYTGTLNDSMNWKEIHQADYLKVEEVVTKCKIKMED 117
129 -----RFIOASYEPLGISLRP-----LEAPPTPTAPPG 159
118 PAFIANPSTSEISITGTELNQOTCLTLRDYNNRKSSEVSTDLIQANPKGALAKSS 177
160 -SPRSEGHDPPTESRSCSQGP-----PSPAPDPKACNMK---KY 197
178 QTKKKKAFNSPKTGQNKTVQYPSDILENASVELFLDANKLPTPVVEQVAQINDSELEL 237
198 KYIVLNSQASQ--AGSLVGERSSGQPCQARLPS---GDEASSSSSSSSSEEGPIPG 252
238 TSVVENTFPAQDVIHTVTKRKKGKQOPNCALKEHSMNSIASVKSPLYEANSGBE--LDQ 295
253 POSRLSPATATVQFCGAPASTPYLLTTSQAODTSGSPSERARP---LPSEFFSCQNC 308
296 RYSKAKPMCNT---CGKVFSEASSLRHRIRHKG-----VKPYVCHLGGKAFQCNQLK 346
309 AVAGCGSLDSLVPGEDDKPYKCOLCRSFRYKGNLASHRTVHTG-EKPYHCSICGARN 367
347 THVRTHTG-----EKPYKCELCDGKFAKCOLVPHSRMHGEEKPKYCDVNLQGA 397
368 RPAHLKTHSRINSGEKPKYKCECTGSRFVQVRSQPPSGFGKPARGGVGQKGFCSGORQ 427
398 TSSNLKIHARKSGEKPKYVCDRCGQRF-----424
428 LKSPSQVAHLRAHVLHTGEEKPYKCECTGSRFRHLQTLKSHVRINHTGKPYHCDPCGLH 487
425 -----AQASTLYTHVRHRTGEEKPYVCDTCGKAFVASSSLITHSRKHTEKPYICGICGS 479
488 FRHKSQRLHLRQKHG 503
480 FISGELNKHFRSHTG 495

```

Search completed: July 14, 2003, 18:14:58  
Job time : 21.9594 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 5.81542 seconds  
(without alignments)  
2620.804 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789  
Sequence: 1 MGSPAPBEGALGVREFTRH.....ROKGAATVTKVHILGSP 518Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfill1.fep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1095.5	39.3	706	1 US-08-074-967-2	Sequence 2, Appl
2	1095.5	39.3	706	2 US-08-553-541B-2	Sequence 2, Appl
3	1095.5	39.3	706	4 US-09-268-202-2	Sequence 2, Appl
4	1095.5	39.3	706	5 PCT-US94-06669-2	Sequence 2, Appl
5	576.5	20.7	803	4 US-09-063-035-2	Sequence 2, Appl
6	419.5	15.0	488	2 US-08-933-750C-17	Sequence 17, Appl
7	419.5	15.0	488	4 US-09-234-613-17	Sequence 17, Appl
8	384.5	13.8	547	1 US-08-340-203A-3	Sequence 3, Appl
9	384.5	13.8	547	2 US-08-452-567-3	Sequence 3, Appl
10	384.5	13.8	547	2 US-08-452-427-3	Sequence 3, Appl
11	384.5	13.8	547	3 US-09-085-407-3	Sequence 3, Appl
12	375.5	13.5	671	3 US-09-121-321-16	Sequence 16, Appl
13	375.5	13.5	671	4 US-08-933-803A-16	Sequence 16, Appl
14	362	13.0	543	4 US-09-362-123A-4	Sequence 4, Appl
15	357.5	12.8	711	2 US-08-820-170A-10	Sequence 10, Appl
16	357.5	12.8	711	4 US-09-055-699-10	Sequence 10, Appl
17	357.5	12.8	711	4 US-09-273-565-10	Sequence 10, Appl
18	357.5	12.8	711	4 US-09-565-538-10	Sequence 10, Appl
19	357.5	12.8	711	4 US-09-561-468-10	Sequence 10, Appl
20	348.5	12.5	462	3 US-08-486-099-117	Sequence 117, App
21	348.5	12.5	462	3 US-08-360-107A-127	Sequence 127, App
22	348.5	12.5	462	3 US-08-484-223B-117	Sequence 117, App
23	348.5	12.5	462	3 US-08-919-597-117	Sequence 117, App
24	348.5	12.5	462	3 US-08-475-668A-117	Sequence 117, App
25	348.5	12.5	462	3 US-08-485-551A-117	Sequence 117, App
26	348.5	12.5	462	3 US-08-471-913A-117	Sequence 117, App
27	348.5	12.5	462	4 US-08-485-264A-117	Sequence 117, App

28	348.5	12.5	462	4 US-08-474-349A-117	Sequence 117, App
29	339	12.2	181	4 US-08-676-318A-44	Sequence 44, Appl
30	329	11.8	181	4 US-08-863-813A-44	Sequence 44, Appl
31	305	10.9	727	2 US-08-475-844-9	Sequence 9, Appl
32	305	10.9	727	5 PCT-US95-08429-9	Sequence 9, Appl
33	305	10.9	728	2 US-08-475-844-5	Sequence 5, Appl
34	305	10.9	728	5 PCT-US95-08429-5	Sequence 5, Appl
35	302	10.8	112	6 5206152-4	Patent No. 5206152
36	299	10.7	338	4 US-09-933-750C-4	Sequence 4, Appl
37	299	10.7	338	4 US-09-234-613-4	Sequence 4, Appl
38	296.5	10.6	648	4 US-09-262-773-2	Sequence 2, Appl
39	296.5	10.6	648	4 US-09-262-773-2	Sequence 2, Appl
40	289	10.4	110	2 US-08-553-541B-9	Sequence 9, Appl
41	289	10.4	110	4 US-09-268-202-9	Sequence 9, Appl
42	281	10.1	106	1 US-08-340-203A-10	Sequence 10, Appl
43	281	10.1	106	2 US-08-452-427-10	Sequence 10, Appl
44	281	10.1	106	3 US-09-085-407-10	Sequence 10, Appl
45	266.5	9.6	429	1 US-08-234-783-4	Sequence 4, Appl

## ALIGNMENTS

## RESULT 1

US-08-074-967-2  
; Sequence 2, Application US/08074967  
; Patent No. 5641672

## GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo

APPLICANT: Chaganti, R.S.K.

TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10112

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/074,967

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/43771

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 706 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-074-967-2

Query Match

Best Local Similarity 35.7%; Pred. No. 3.4e-76;

Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

QY 1 MGSPAPBEGALGVREFTRHSDVGNLRLRGLTDVTLVGGCPRLPARKATLACS 60

DB 1 MASPA--DSCI---QFTRHARDVLLNRLRSRDLTDVIVVSRQFRAKTVLAWMR 54

QY 61 GFFYSIFRAGVGVVSLPGPEARGFAPLLDFMYTSLRLSPATAPAVLAATYLOM 120  
 DB 55 GLFYSTFDQKCNLSVINDPEINPEGFCILDFMYTSLRLREGIMAVMATYLOM 114  
 QY 121 EHVVOACHRTIOAS----- 134  
 DB 115 EHVVDTCRKFIKASEAMVSAIKPREEFNLNMLPQDIMAYRGVEVNNLPLRSAPG 174  
 QY 135 -----YEPGLGIS----- 141  
 DB 175 CESRAFPAPSLVGLSTPPASYSMSHLVPSLLFSDDEFRDVMPVAMPFKERALPCDS 234  
 QY 142 -----LRPL----- 145  
 DB 235 ARPVEGYSRPTLEVPNVCHSNISPKETIPEARSMDHYSVABGLKPAAPASARNAPYF 294  
 QY 146 -----EAPPT-----PPTAP-----PPGSRREBGPDPPTES-----RSC 177  
 DB 295 PCDKASKEERPSSEDEIALHFEPPNAPLNRKGLVSPQSKDCQPNSPTEACSSKNAC 354  
 QY 178 -----SOGPPSPASPPKACNMKKYKYIVLNS-----QASQAGSLVGRSSGQPCP 223  
 DB 355 ILQSGSPPAKSPPTDPKACSMKKYKFLVNSLNONAKPGPEQAEIGRLSPRAYTAPAC 414  
 QY 224 QARLPBGDEASSSSSSSSSEEGPIPGPOSRL-----SPTATVGF----- 266  
 DB 415 QPFW-BEENDLQSPTKLSASGEDSTIP-QASRLNNIVNRSMTGSPRSSSESHPLYMHP 472  
 QY 267 -----KCG--APASTPYLTSQ-----ADTSGSPSERAPLPGESEFSSCONCEAVAGCSG 316  
 DB 473 PKCTSGSGSPQAHAEMCLHTAGPTFAEMGETQSEYSSSCENGAFCECCDRFSEBAS 532  
 QY 317 LD-SLVPGDEDEKPYKCOLCRSFRYKGNLASHRTVHTGKPYHCISICGARFNPANLKTTH 375  
 DB 533 LKHTIQTHTDCKRYKCDRCQASFRYKGNLASHRTVHTGKPYHCINICGAQFNPANLKTTH 552  
 QY 376 SRHSGEKPKYKCTCGSRFVQVRSQPPSGFGKPARGGVGQKGCSSQRDLKSPPOV 435  
 DB 593 TRHSGEKPKYKCTCGARFV-----QY 614  
 QY 436 AHIRAVLHTGKPYKPCPTCGTRFRHLQTLKSHVRIHTGKPYHCDPGLHFRHKSQLR 495  
 DB 615 AHIRAVLHTGKPYKPCPTCGTRFRHLQTLKSHVRIHTGKPYHCEKCNLHFRHKSQLR 674  
 QY 496 LHLRQKGAATNTKVHYHI 514  
 DB 675 LHLRQKGAATNTKVQYRV 693

RESULT 2  
 US-08-553-541B-2  
 Sequence 2, Application US/08553541B  
 Patent No. 5882858  
 GENERAL INFORMATION:  
 APPLICANT: Datta-Favera, Riccardo  
 APPLICANT: Chaganli, Raju S.K.  
 TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS  
 TITLE OF INVENTION: bcl-6  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: United States of America  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Releasee #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/553,541B

FILING DATE: May 28, 1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 0575/43771-A-PC-T-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 TELEX: 422523 COOP UI  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 706 amino acids  
 TYPE: amino acid  
 TOPOLOGY: 1linear  
 MOLECULE TYPE: protein  
 US-08-553-541B-2

Query Match 39.3%; Score 1095.5; DB 2; Length 706;  
 Best Local Similarity 35.7%; Pred. No. 3.4e-76;  
 Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

QY 1 KSPAPBAGALGYBEFTRHSDVYGNLNEFLRGLIDVTLLVGGQPLRAKAVLIACS 60  
 DB 1 MASPA--DSCI-----QFTHARDVLLNLRSLRSDILTDVIVVSGEPPRAKTYLMAWR 54  
 QY 61 GFFYSIFRAGVGVVSLPGPEARGFAPLLDFMYTSLRLSPATAPAVLAATYLOM 120  
 DB 55 GLFYSTFDQKCNLSVINDPEINPEGFCILDFMYTSLRLREGIMAVMATYLOM 114  
 QY 121 EHVVOACHRTIOAS----- 134  
 DB 115 EHVVDTCRKFIKASEAMVSAIKPREEFNLNMLPQDIMAYRGVEVNNLPLRSAPG 174  
 QY 135 -----YEPGLGIS----- 141  
 DB 175 CESRAFPAPSLVGLSTPPASYSMSHLVPSLLFSDDEFRDVMPVAMPFKERALPCDS 234  
 QY 142 -----LRPL----- 145  
 DB 235 ARPVEGYSRPTLEVPNVCHSNISPKETIPEARSMDHYSVABGLKPAAPASARNAPYF 294  
 QY 146 -----EAPPT-----PPTAP-----PPGSRREBGPDPPTES-----RSC 177  
 DB 295 PCDKASKEERPSSEDEIALHFEPPNAPLNRKGLVSPQSKDCQPNSPTEACSSKNAC 354  
 QY 178 -----SOGPPSPASPPKACNMKKYKYIVLNS-----QASQAGSLVGRSSGQPCP 223  
 DB 355 ILQSGSPPAKSPPTDPKACSMKKYKFLVNSLNONAKPGPEQAEIGRLSPRAYTAPAC 414  
 QY 224 QARLPBGDEASSSSSSSSSEEGPIPGPOSRL-----SPTATVGF----- 266  
 DB 415 QPFW-BEENDLQSPTKLSASGEDSTIP-QASRLNNIVNRSMTGSPRSSSESHPLYMHP 472  
 QY 267 -----KCG--APASTPYLTSQ-----ADTSGSPSERAPLPGESEFSSCONCEAVAGCSG 316  
 DB 473 PKCTSGSGSPQAHAEMCLHTAGPTFAEMGETQSEYSSSCENGAFCECCDRFSEBAS 532  
 QY 317 LD-SLVPGDEDEKPYKCOLCRSFRYKGNLASHRTVHTGKPYHCISICGARFNPANLKTTH 375  
 DB 533 LKHTIQTHTDCKRYKCDRCQASFRYKGNLASHRTVHTGKPYHCINICGAQFNPANLKTTH 592  
 QY 376 SRHSGEKPKYKCTCGSRFVQVRSQPPSGFGKPARGGVGQKGCSSQRDLKSPPOV 435  
 DB 593 TRHSGEKPKYKCTCGARFV-----QY 614  
 QY 436 AHIRAVLHTGKPYKPCPTCGTRFRHLQTLKSHVRIHTGKPYHCDPGLHFRHKSQLR 495  
 DB 615 AHIRAVLHTGKPYKPCPTCGTRFRHLQTLKSHVRIHTGKPYHCEKCNLHFRHKSQLR 674  
 QY 496 LHLRQKGAATNTKVHYHI 514  
 DB 675 LHLRQKGAATNTKVQYRV 693



## RESULT 3

US-09-268-202-2  
Sequence 2, Application US/09268202

Patent No. 6174997

GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo

APPLICANT: Chaganti, Raju S.K.

TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS

TITLE OF INVENTION: bcl-6

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/268,202

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US-Z

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

TELEX: 422523 COOP UT

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 706 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-268-202-2

Query Match 39.3%; Score 1095.5; DB 4; Length 706;

Best Local Similarity 35.7%; Pred. No. 3.4e-76; Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

1 MGSPAPGALGYRBFTRHSDVLDGNLNLRLGLITDVTLLVGGQPLRAHKAVALIACS 60  
1 MASPA--DSCI--QFTRHARDVLLNLRLRGRDILTDVIVVSRGQFRAHKTIVLMAVR 54  
61 GFFYSIFRGRAGVGVVLSLPGCEPGRAPLIDFMYTSRLRLSPATAVAVALAATYLLM 120  
55 GLFYSIFRGRAGVGVVLSLPGCEPGRAPLIDFMYTSRLRLSPATAVAVALAATYLLM 114  
121 EHVYQACHRFIOAS-----YEPGLGIS----- 134  
115 EHVYDTCRFTKASEAMVSAIKPPEBPLNSMLMPODIMAYRGREVENNLLPLSARG 174  
135 -----YEPGLGIS----- 141  
175 CESRAPAPSLYGLSTPPASYSWYSHLPVSSLLFSDBEPRIVMVPVAPPEKRALPCOS 234  
142 -----LRPL----- 145  
235 APPVPGESRPTLEVPNVCHSNISYPKETIPREASDMHTSYAEGIKPAAPARNAPYF 294  
146 -----EASPT-----PTAP-----PGSPRRSGHPDPTES-----RSC 177  
295 PCOKAKERERPSSEBIALHFPFPAALNRKGLVSPQSTQKSDCQPNSTETECSSKMAC 354  
178 ----SGPPSPASPDPKACNMKTKYKIVLNS-----QASQAGSLVGRSSGQPCP 223

DB 355 ILQGGSPSPAKSPTPDKACSMWKYKIVLNSLNQNAKCGPEQALGRLSPRAYTAPAC 414  
224 QARLPSGDEASSSSSSSSSEEGPIPOPOSRL-----SPTATVGF----- 266  
DB 415 QPPM-EPENLDQSTKXASAGEDSTIP-QASRLNNIVNRSMTSPPRSSSEHSPLYMHP 472  
267 ----KCG--APASTPYLLTSQ----ADPTSGSPSERAPPLPGSEFFSCONCEAVAGCSG 316  
DB 473 PKCTCGSGSPQHAEMCHTAGPTFAEMWGTQSEYSSSCENGAFFCNBCDCRRSEAS 532  
317 LD-SLVPGDEDEPKYQCLCRSSFRYKGNLASHRTYHTGKPHGSCICGARFRRPANKTH 375  
DB 533 LRRHTLQTHSDPKYCDRCQASFRYKGNLASHRTYHTGKPHGSCICGARFRRPANKTH 592  
376 SRHSGEKPKYKCTGSRFVQVRSQPPSGFGKPARGVGQKGFCSQRDLKSPPOV 435  
DB 593 TRHSGEKPKYKCTGSRFVQVRSQPPSGFGKPARGVGQKGFCSQRDLKSPPOV 614  
436 AHLRAHVLHTGKEKYPCTGTRPRHLQTLKSHRHTGKPYACDPGLHFRKSOQR 495  
DB 615 AHLRAHVLHTGKEKYPCTGTRPRHLQTLKSHRHTGKPYACDPGLHFRKSOQR 674  
496 LHLRQHGATNTKYHYHI 514  
DB 675 LHLRQHGATNTKYHYHY 693

## RESULT 4

PCT-US94-06669-2  
Sequence 2, Application PC/TUS9406669

GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo

APPLICANT: Chaganti, R.S.K.

TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS

TITLE OF INVENTION: bcl-6

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/06669

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/074,967

FILING DATE: 09-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 43771-A-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UT

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 706 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-06669-2

Query Match 39.3%; Score 1095.5; DB 5; Length 706;

Best Local Similarity 35.7%; Pred. No. 34e-76;  
Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18

QY	1	MGSPAPBEGALGVYRREFTSHSSVDVGNLNLNLTARGLTIDVTYTLVVGQCPRLBAHRAVLIACS	60
Db	1	MSPPA--DSCI-----OFTRHARDVLLNLNLRSDRLITDVIVVSRQFPAHKTIVLMAWR	54
QY	61	GEFYISIFRGRAGVGVULSLPGGPEARGFAPLIDFMYTSRLRLSPAPAPVLAAATYLOM	120
Db	55	GLFYISFTQLOKCNLSVINUDPINEGFCIILDDPMYTSNLRLREGIMAVMATYLOM	114
QY	121	BHVVOACHRPLOAS-----	134
Db	115	BHVVDTCRKFIXASEAMVSAIKPRBEPFLNSMLMPQDIMAYRGVEVENNLPLRSAPG	174
QY	135	-----YEDLGIS-----	141
Db	175	CESRAFAPLVGLSTPPASYSMTSHLPVSSLLPSEDEFDVMPPANPFPKERALECDS	234
QY	142	-----LRPL-----	145
Db	235	ARPVGEYSRPTLEVS PNVCHSNIVPKETIPEASSDMHYSVAEGLKPAAPBARNAPYF	294
QY	146	-----EAEPT-----PTAP-----PPGSRREGEHPDPTES-----RSC	177
Db	295	PCDKASKKEERSESSDEELAHFEPNAPLNRKGLVSPQSPQSDCQNSPTKACSSKNAC	354
QY	178	-----SGSPSPASPDCKCNMKKKYKTVINS-----QASQAGSLVGRHSSGQPCP	223
Db	355	IIQSGSPSPAKKPTDPRKACSWKKYKTVINSLNQNAKPGGEQAEGLRSLPRAVYAPAC	414
QY	224	QARLPSEGDASSSSSSSSSSSEEGIPGQSRU-----SPTAAVQF-----	266
Db	415	QPPM--EPENLUDIOSPTKLSASGDSITP--QASLNINI VNRSMGTSPRSSSEHSPLYMHP	472
QY	267	---KCG--APASTPYLLTSQ---AODTSGPSEPARPLPGSEFPSCONCEAVAGCSSG	316
Db	473	PKCTSGSGSQSPQHAEMECMLHTAGPTFAEMEGTOSBYSDSSCENGAFECNCDRFSSEAS	532
QY	317	LD-SLVPGDEDPYKQCOLCRSSPRYAGNLASHHTVHTGEXPYCSTCGAEFNRPAULKTH	375
Db	533	LGRHLLQTHSDKPYKCDKRCOSFRYVGNLASHHTVHTGEXPYCNCICGAOFNRPANLKTH	592
QY	376	SRHSGEKPYKGETCGSRFVQVVSOPPSGFGQKPARAGVGOQKGAFCSQRODLKSPSPQV	435
Db	593	THHSGEKPYKGETCGARV-----QV	614
QY	436	AHLRAHVLITHGEKPYPCPTCGTRFPHLOTLKSHVRIHTGEKPYHCDPCGLAFHRSQLR	495
Db	615	AHLRAHVLITHGEKPYPCPEICGTRFPHLOTLKSHRIHTGEKPYHCEKCNLAFHRSQLR	674
QY	496	LHLROKGAATNTKYAHYHI	514
Db	675	LHLROKGAATNTKYOYRV	693

RESULT 5  
US-09-063-035-2  
: Sequence 2, Application US/09063035

```

? GENERAL INFORMATION:
? APPLICANT: PEUKERT, Karen; HANEL, Frank; and EILERS
? APPLICANT: Martin
? TITLE OF INVENTION: Myc-binding zinc finger proteins
? TITLE OF INVENTION: their preparation and their use
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
?

```

COMPUTER READABLE FORM:

**MEDIUM TYPE:** Diskette, 3.5 inch, 1.2 Mb storage  
**COMPUTER:** IBM AT-compatible, 80486 processor  
**OPERATING SYSTEM:** MS-DOS version 6.1

```
SOFTWARE: WordPerfect version 8.0
CURRENT APPLICATION DATA:
```

AFFILIATION NUMBER: 05/03  
 FILING DATE: 21-APR-1998  
 CLASSIFICATION: 514  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:

US-09-063-035-2

Query Match	20.7%;	Score 576.5;	DB 4;	Length 803;
Best Local Similarly	23.4%;	Pred. No. 2.8e-36;		
Matches 180;	Conservative 64;	Mismatches 205;	Indels 163;	Gaps 20

Qy	16	EFRRHSADVLMNLTNLTGLITDVTLVVGQPLRANHAVLIASGFFYIIFRRAG-VG	74
Dd	2	DFQSHQVLTBEQNLQGRGLGCLDCTFYVDGVNHRKAKAVLAASSEYFKYLFYDQKDVH	61
Qy	75	VDVLSLPGREARGFAPLLDEMYTSLRLSPATAPAVLAATYUOMENVVQACHRTQAS	134
Dd	62	LDI-----SNAAGLGOMLEFMYTAKTSLSPENVDVLAATVFLQMODITACH-----	109
Qy	135	YEBLGISLRPLAEPRTP-----TAPPSPPRSSEHPRPRPSSSSCQGPSPASPD	188
Dd	110	-----ALXSL-ABEPTSPGNAELATEGDDKAKEBKATSTLSLBEQAGSTPGP-	161
Qy	189	PKACNNKKYKYIVLNSQASQAGSLVGRSSGQPCPOARLP-----SGDEASSSSSSSSSS	245
Dd	162	-----SRDLKERGGQQAASAGAQTEKADAPREPPVELKDPRTSGMAAABABALSSSE	219
Qy	246	EEBPT-PEGQSLSPATVQPKCA-PAST-----PYLITSQA-----QDTGSPF--	290
Dd	220	QEMEVVPARKGEEQKEQEBEGBEAGPAEYKEGSOLENGAEPEENENESAGTSGQE	279
Qy	291	--ERAPLP-----GSSF-----FSCONC	307
Dd	280	LGSEANGLASGYGDRTESKAVGSYIHKCEQCGEFYHTNGFKHRIIHTGKRFSCREC	339
Qy	308	EAV-----ACSSGGLDVLVGEDKPKYQOLCRSSFRY-----	340
Dd	340	SKAFSPDAPACKAHEKTHSP---LKYGCEGCGSKGYSRLISLNLHKRHSGEARYRCEDCG	366
Qy	341	-----KGNLASRHYHTGKPYHNCIG-----AREN	367
Dd	397	KLETTSGNKKRIQLVHSGEKPYQDCVCGRSFSDPTSKRHLETHDTEKHKCPHCDKFN	456
Qy	368	RPANLKTSHRSIHSGEKPYCETGSGRFPQVNSOPRPSGFGQKPAAGVGQKGFSSQROD	427
Dd	457	QVGNLKAHLKIHLADGRLKRCGCGQFL-----TSGNLKROLRIHSGEKPYCINHQRO	510
Qy	428	LKSPBSQVAAHRAVLIHTGKPYCPTGCGTRFPHLOTLKSHVRIHTGKPYHNCPCGLH	487
Dd	511	FADPGA-----LQRHRIHTGKPYCQCVWACGAFTQASLLIHYQHHTGKPYCERCGKR	566
Qy	488	FRHKSQLRLHR	499
Dd	567	FVQSSQLANHR	578

RESULT 6  
US-08-933-750C-17  
; Sequence 17, Application US/08933750C

```

; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga

```

APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,750C  
FILING DATE: September 23, 1997  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRISTNOT03  
CLONE: 641127  
US-08-933-750C-17

Query Match 15.0%; Score 419.5; DB 2; Length 488;  
Best Local Similarity 29.5%; Pred. No. 1.7e-24;  
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

176 SCSGPPSPASPDPKACN-----WKYKYIYVLSOA-----SQAGSLVGRSS 218  
137 TCQNG-QLKESLDPIQCKDKIHGK--SQVSCSQQRGHTBEKPCDHNNCGKIILNTSPD 193  
219 GPCPCOARLPSCGDEASSSSSSSSSEGGPIPGQSRUSPTAATVQFKGAPASTPYLL 278  
194 GHPYEKIHAEKQYSGSCGKNFSQSS-----LL 223  
279 TSOAODTSSPSRRARPLPGESEFFSCQNC-EAVAGCSSGLDSLVPDDEKPYKCOLCRSS 337  
224 LHQRDHTKEP-----YKCEQCGKGFTRSSSLIHQAVHTDEKPYKCDKCGKG 271  
338 FRYKGNLASHRTVHTGEKPYHCSICGARFNRPAHLKTHSHIGSEKPYKCEGSRFVQ- 396  
272 FTRSSSLILHNAVHTGEKPYKCDKCGKFSQSKLTHQVHTGEKPYCEGCEGMSFSQR 331  
397 ---VRSQPSGFGKPARAGVGQKCGFSSQRQDLKSPSQVAHLRAVHLHTGKPYR 452  
332 SNLHIGRYHTG--EPRPYKCGECGKF-----SQSSNLHIRHCIRHTGKPYQ 376  
453 CPTGGRFPHLQTLKSHVRIHTGEKPYHCDPQGLHFRHKSQ-----RLHLRQK----- 501  
377 CTECGGFSQSSDLRLHLRVHTGEKPYHCGKCGKFSQSSKLHIGRVHTGEKPYKCDKCGK 436

QY 502 -HGAANTKRVHYH 513  
DB 437 GKGFSQSSNLHITH 449

RESULT 7  
US-09-234-613-17  
Sequence 17, Application US/09234613  
Patent No. 6132973

GENERAL INFORMATION:

APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/234,613

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750

FILING DATE: September 23, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 488 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BRISTNOT03

CLONE: 641127

US-09-234-613-17

Query Match 15.0%; Score 419.5; DB 4; Length 488;  
Best Local Similarity 29.5%; Pred. No. 1.7e-24;  
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

176 SCSGPPSPASPDPKACN-----WKYKYIYVLSOA-----SQAGSLVGRSS 218  
137 TCQNG-QLKESLDPIQCKDKIHGK--SQVSCSQQRGHTBEKPCDHNNCGKIILNTSPD 193  
219 GPCPCOARLPSCGDEASSSSSSSSSEGGPIPGQSRUSPTAATVQFKGAPASTPYLL 278  
194 GHPYEKIHAEKQYSGSCGKNFSQSS-----LL 223  
279 TSOAODTSSPSRRARPLPGESEFFSCQNC-EAVAGCSSGLDSLVPDDEKPYKCOLCRSS 337  
224 LHQRDHTKEP-----YKCEQCGKGFTRSSSLIHQAVHTDEKPYKCDKCGKG 271

Qy	338	FRYKNLSHRTVHNGEKPYPHGSICGAPRNRPAHLKTSRHSIGKEPKYCEGSRFPVQ	396
Dd	212	FTRSSLLIHAHVHTEGKEPKYCDCKGKGFSSQSKLIIHQRVHTEGKEPKYCECGMSFSOR	331
Qy	397	---VRSQPPSGFOGKPARGVGVQKGFSSQSRDLKSPSOVAHLRAVLIHTGKEPKY	452
Dd	332	SNLHIHQVHTC--ERPXYKCGECGKG-----SQSSNLHIHRCIHTGKEPKYQ	376
Qy	453	CPTGTRRRLIOTLKSVAIRIHTEGKEPKYCDPCGLFPRHKSQV-----RLHLRQK-----	501
Dd	377	CYECKGKGFSSQSDLIHLRVHTEGKEPKYHCKGKGFSSQSKLIIHQRVHTEGKEPKYCECSK	436
Qy	502	-HGAATNTKVVHY 513	
Dd	437	GKGFSSQSSNLHIH 449	
<p>RESULT 8</p> <p>US-08-340-203A-3</p> <p>Sequence 3, Application US/08340203A</p> <p>Patent No. 5756668</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Baylin, Stephen B.</p> <p>APPLICANT: Walee, Michele M.</p> <p>TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1</p> <p>NUMBER OF SEQUENCES: 14</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Fish &amp; Richardson P. C.</p> <p>STREET: 4225 Executive Square, Suite 1400</p> <p>CITY: La Jolla</p> <p>STATE: California</p> <p>COUNTRY: USA</p> <p>ZIP: 92037</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent In Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/340, 203A</p> <p>FILING DATE: 15-NOV-1994</p> <p>CLASSIFICATION: 530</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Haile, Ph.D., Lisa A.</p> <p>REGISTRATION NUMBER: P-38,347</p> <p>REFERENCE/DOCKET NUMBER: 07265/039001</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (619) 455-5100</p> <p>TELEFAX: (619) 455-5110</p> <p>INFORMATION FOR SEQ ID NO: 3:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 547 amino acids</p> <p>TYPE: amino acid</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: protein</p> <p>US-08-340-203A-3</p>			
Qy	Query Match	13.8%; Score 384.5; DB 1; Length 547;	
Dd	Beet Local Similarity	24.6%; Pred. No. 9,4e-22;	
	Matches 128; Conservative 54; Mismatches 170; Indels 169; Gaps 17		
Qy	20	HSDDVLGNLRLRLGILTDVTLVGGQPLRAHKAVLIJACSGFYSIFRGRAGVGDVLS	79
Dd	10	HSRQLLDLNNORTKGFGLCDVIIVQNALFPAHKAVLAASSAYLSLV-----VNDNLIN	64
Qy	80	LPGGEARG-FAPLLDFMYTSRL-----RLSPATAP-----AVLAATYLGMEHVQ	125
Dd	65	LDHDMVSPAVFRVLVDFTYTGRLADGAEMAAVAAPGAEPELGAVLAAASYLGIPDLVA	124
Qy	126	ACHRFIQ-----ASTEP-----LGILRLPLEAEPPTTPPAAP-FGSPR	162
Dd	125	LCKRLKHKGVYCHLRGGGGGGGYAPYAMTSMAGSAAPASAKSVVGTTPSRPGPR	184

```

QY 163 RS-----EGHPDPTSSRSOSGPPSPASDPACMVKKYKYLIVLSQASQA 209
DB 185 SAMRRRATTLAAMTGPARPARATATTTTAAARFVAARPAAP----- 225
QY 210 GSLGERSGCPCPQAR-----LPSGDEASSSSSSSS--SSSSECGPIPG 253
DB 226 ----GHNLEGVPCPHLAVGPEPESFGDNLVYCIPEGKGPSPSEQLNAVHVAEE----- 275
QY 254 QSRLSPTNATVQFCGAPASTFPYLLTSGAODTSGSPSERAPPLPGSEFFSCQNEAVAGC 313
DB 276 BEALYGRAEAEEVAAGAAGLAP-----PFGGGG-----DKVAGA 309
QY 314 SSGGLDSLVPEDEDPEPYCQQLCRSSFFRYKGNLASRTVHTGEXPHCSICGARFNRPAVLK 373
DB 310 PGGLGELI-----RPIGSGSDKSKYKOPATIRQEKHTHMLTRPYPCTICGKKTFQRCMT 364
QY 374 THSRHSGEKPKYKCEGCSRFVQVRSQPPSGFGCKPARGGVGOKGFCGSSQRODLKSPPS 433
DB 365 RHMRSHGLKRFACDACGMRTTR----- 387
QY 434 QVAHLRAVLII-TGEKYPPEPTGCTFRHLQTLKSHYRIH 473
DB 388 QYRLTRTHRIHPREKPKYECQVCGKRFADQRLNLSHKWH 428

RESULT 9
US-08-452-567-3
; Sequence 3, Application US/08452567
; Patent No. 5846712
; GENERAL INFORMATION:
; APPLICANT: Baylin, Stephen B.
; APPLICANT: Wales, Michele M.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,567
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/039003 (Div PD3664)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-567-3

Query Match 13.8%; Score 384.5; DB 2; Length 547;
Best Local Similarity 24.6%; Pred. No. 9,4e-22;
Matches 128; Conservative 54; Mismatches 170; Indels 169; Gaps 17;

```

```

QY      80  LFGPEARG-FAPLLDPMYTSRL-----RLSPATAP---AVLAATYLQMEHVQ 125
      65  LDHDMVSPAVFRLVDFIYTGRLADGAEAAAAAVALPGAEPLIGAVLAASYLQIPDLVA 124
QY      126  ACHRIQ-----ASYEP-----LGISLRLAEPPPTPTAP--PGSPR 162
      125  LCKKRLKRGKXCHLIRGGGGGGGYAPYAMATSWAGSAAPPAKSAKSVGTSPSPGGR 184
QY      153  RS-----EGHPDPPTESRSCSQGPPSPASPDPKACNMKKYKYLINSQASQA 209
      155  SAMRRRRATLAAWTGPAAATATTTRAAARRVAAATPAP----- 225
QY      210  GSLVERSSGQPCPOAR-----LPSGDEASSSSSSSS--SSSSEGP1PGP 253
      226  ----CGHLEGYPCPHLAYGEPESFGDNLVVCIPCGKFPSSBQLNAHVAHYE----- 275
QY      254  QSRLEPTAATVQFKGAPASTPYLLTQADTSSGSEBARPLPSSEFPSCONCEAVAGC 313
      276  BEALYGRAEAAVAAGAGLGP-----PFGGG-----DKVAGA 309
QY      314  SSGLDLVPGDEBDKPYKCOLCRSPRYKGNLASHRTVHTGKPYHCSICGAEFNPANLK 373
      310  PGGLGELL-----RPRGSCDCKSYKDPATLROHEKTHWLTTPYCTICGKFTQGRMT 364
QY      374  THSRHSGEKPYKCTCGSRFVQVRSQPPSGFGKPARGVGQKGFSSORQDLKSPS 433
      365  RMRSHLGLKPPACACGMRFT----- 387
QY      434  QVAHLRAVYLH-TGEKPYPCPTCGTRPHLOTLSHYRIH 473
      388  QYRLTRTHRIHPRGEKPYECVCGKFAQORNLISHMKH 428
Db

```

## RESULT 10

```

US-08-452-427-3
; Sequence 3, Application US/08452427
; Patent No. 5922590
; GENERAL INFORMATION:
; APPLICANT: Baylin, Stephen B.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,427
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; PRIORITY APPLICATION NUMBER: US 08/340,203
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REFERENCE/DOCKET NUMBER: 07265/039001
; TELEPHONE: (619) 455-5110
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-427-3

```

Query Match 13.8%; Score 384.5; DB 2; Length 547;

Best Local Similarity 24.6%; Pred.No. 9,4e-22;

Matches 128; Conservative 54; Mismatches 170; Indels 169; Gaps 17;

```

QY      20  HSDVILGNINELRLGILTDVTLVGGQPLRAHKAVALJACSGFPFIFRGRAGVGDVLS 79
      10  HSRQLDLQNNQRTGFLCDVIIVQNALFRHKRVLAASSAYLSLV-----VINDNLN 64
Db
QY      80  LFGPEARG-FAPLLDPMYTSRL-----RLSPATAP---AVLAATYLQMEHVQ 125
      65  LDHDMVSPAVFRLVDFIYTGRLADGAEAAAAAVALPGAEPLIGAVLAASYLQIPDLVA 124
QY      126  ACHRIQ-----ASYEP-----LGISLRLAEPPPTPTAP--PGSPR 162
      125  LCKKRLKRGKXCHLIRGGGGGGGYAPYAMATSWAGSAAPPAKSAKSVGTSPSPGGR 184
QY      153  RS-----EGHPDPPTESRSCSQGPPSPASPDPKACNMKKYKYLINSQASQA 209
      155  SAMRRRRATLAAWTGPAAATATTTRAAARRVAAATPAP----- 225
QY      210  GSLVERSSGQPCPOAR-----LPSGDEASSSSSSSS--SSSSEGP1PGP 253
      226  ----CGHLEGYPCPHLAYGEPESFGDNLVVCIPCGKFPSSBQLNAHVAHYE----- 275
QY      254  QSRLEPTAATVQFKGAPASTPYLLTQADTSSGSEBARPLPSSEFPSCONCEAVAGC 313
      276  BEALYGRAEAAVAAGAGLGP-----PFGGG-----DKVAGA 309
QY      314  SSGLDLVPGDEBDKPYKCOLCRSPRYKGNLASHRTVHTGKPYHCSICGAEFNPANLK 373
      310  PGGLGELL-----RPRGSCDCKSYKDPATLROHEKTHWLTTPYCTICGKFTQGRMT 364
QY      374  THSRHSGEKPYKCTCGSRFVQVRSQPPSGFGKPARGVGQKGFSSORQDLKSPS 433
      365  RMRSHLGLKPPACACGMRFT----- 387
QY      434  QVAHLRAVYLH-TGEKPYPCPTCGTRPHLOTLSHYRIH 473
      388  QYRLTRTHRIHPRGEKPYECVCGKFAQORNLISHMKH 428
Db

```

## RESULT 11

```

US-09-085-407-3
; Sequence 3, Application US/09085407
; Patent No. 6103877
; GENERAL INFORMATION:
; APPLICANT: Baylin, Stephen B.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,407
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; PRIORITY APPLICATION NUMBER:
; FILING DATE: 15-NOV-1994

```

```

; APPLICANT: Baylin, Stephen B.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,427
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; PRIORITY APPLICATION NUMBER: US 08/340,203
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REFERENCE/DOCKET NUMBER: 07265/039001
; TELEPHONE: (619) 455-5110
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 amino acids
; TYPE: amino acid

```

```

?      ATTORNEY/AGENT INFORMATION:
?      NAME: Haile, Ph.D., Lisa A.
?      REGISTRATION NUMBER: P-38,347
?      REFERENCE/DOCKET NUMBER: 07265/039000
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (619) 455-5100
?      TELEFAX: (619) 455-5110
?      INFORMATION FOR SEO ID NO: 3:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 547 amino acids
?      TYPE: amino acid
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
US-09-085-407-3

```

Query Match	13.8%	Score 384.5;	DB 3;	Length 547;
Best Local Similarity	24.6%;	Pred. No. 9,4e-22;		
Matches 128;	Conservative 54;	Mismatches 170;	Indels 169;	Gaps 17;

```

Qy 20 HSSVJGNLNELELRGLDVTLLVGGCPLRAHKVLLACSGFFSIFIRGAGVGVLS 79
Db 10 HSNQQLLQANNQTKGFLCDVIIYVONALFPHKIVLAASSAYLKSIV-----VHDLNL 64
Qy 80 LPGGPEARG-FAPLDPMYTSRI-----RUSPATAP---AVLAATYLOMEHVQ 125
Db 65 LDHDMVSPAVFRLVDFIYTGRLADGAEAAAAAVALPGAEPLGAVLLAAAYLOIDPLVA 124
Qy 126 ACHRFIO-----ASYEP-----LGISLRPLEAEPPTPTPAAP-POSPR 162
Db 125 LCKKRLKHGKYCHLAGGGGGGGAIFYAMATSMGSAAPPANSAKSVGTRPSRPGGPR 184
Qy 163 RS-----EGHDPPTESHSCSGGPPSPADPKACMVKYIVLNSQASQA 209
Db 185 SAMRRRATLAATGPARAATAATTTAAARRRVAARTAPAP-----225
Qy 210 GSVLGERSSGQPCPOAR-----LPQDDEASSSSSSSS---SSSSEGP1RG 253
Db 226 ----GHLBEGYPCPHLAYGPESFGNLVYC1PCGKGFPSSQULNAHYEAIVE-----275
Qy 254 QSRUSPTAATVQPKCAPASTPYLLTSAQDPTSGSPSERARLPGESEFFSCONCEAVAGC 313
Db 276 EEAALYGRAAEAEEVAAAGALGP-----PFGGG------DKVAGA 309
Qy 314 SSGGLDLYSPEDDEKPYKCOLCRSSPFRYKGNLSHRTVHTGEPKPYCS1CGARFMRPNLK 373
Db 310 PGLIGELL-----RPRCGSCDKSYDPAITLQHKHTMLTRPCTICGKFK1QRGTM 364
Qy 374 THSRHSGEKPYKCYCETCGSRFVQVRSQPSGCGQPARGAVGQKGFCSQRODLKSPS 433
Db 365 RHRMRSHGLKRFACDACGMKFR-----387
Qy 434 QVALRAHVLIH--TGEKPYPCPTCGTRFRHLDTLKSRIH 473
Db 388 QYRLTRTHMKIHRGEPKPYCOVCGGKFKQOQNNLLSHMKH 428

```

RESULT 12 321-16  
US-09-121-321-16  
Sequence 16, Application US/09121321  
Patent No. 6090783  
GENERAL INFORMATION:  
APPLICANT: Saiga, Akihiko  
APPLICANT: Orica, Satoshi  
APPLICANT: Igarashi, Hisanaga  
APPLICANT: Okumura, Kouichi  
APPLICANT: Sakaguchi, Gaku  
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF  
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York

```

1 STATE: New York
2 COUNTRY: USA
3 ZIP: 10020
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patentin Release #1.0, Version #1.3Q
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/09/121,321
12

```

Query Match	13.5%	Score 375.5	DB 3	Length 671
Best Local Similarity	28.3%	Pred. No. 66-21		
Matches 106	Conservative 26	Mismatches 114	Indels 129	Gaps 15

[illegible]

RESULT 13  
 US-08-933-603A-16  
 Sequence 16, Application US/08933803A  
 Patent No. 6218522  
 GENERAL INFORMATION:  
 APPLICANT: Saiga, Akiniko  
 APPLICANT: Orita, Satoshi  
 APPLICANT: Igarashi, Hisanaga  
 APPLICANT: Okumura, Kouichi  
 APPLICANT: Sakaguchi, Gaku

TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF  
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,803A  
FILING DATE: 19-SEP-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: SHGN-12CIP  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 671 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-933-803A-16

Query Match 13.5%; Score 375.5; DB 4; Length 671;  
Best Local Similarity 28.3%; Pred. No. 6e-21;  
Matches 106; Conservative 26; Mismatches 114; Indels 129; Gaps 15;  
QY 152 PPTAPPSPGPRRSEGHDPPTESRSCSG-PP-----SPASPDPKACMKKYYIYLSNQ 205  
DB 403 PPPAPPPOPOPOPO-POPOLOSOPPOPSLPPIAVALNPGGPPSRG-----LLND 450  
QY 206 ASQASLVGERSGQPCPOARLPSCGDEASSSSSSSSSSSSSEKPIPTGQSRLSPTAATVQ 265  
DB 451 GQV---LPERGSGE-AP-----PGGRSTGGGGGCGGGGG----- 484  
QY 266 FKGCAAPASTFYLLTQAQDTSGSPSERAPPLPGESEFSCQCEAVAGCSSGL--DSLVPG 323  
DB 485 ---GAEAGT-----GAGG-----GCGSC---CPGGLRRSLLLHG 512  
QY 324 DEDKPYKQCLGSSPFYKGNLASHRTVHTGKPYHCSIGAPFNRPNALKTSRIHSGEK 363  
DB 513 ASKPTSCPECGSKSFGVRKSLIITHRSHTKERPYEACBCEKSKNCHSLIRHOMTRGER 572  
QY 384 PYKCTCGSRFYQVRSQPPSGFGKPARGVGQKGFCSJDRDLKSPSOVAHLFAVYL 443  
DB 573 PYKCSCEKTY-----SRKEHLQNHQR 594  
QY 444 IHTGEPYPCPTGCTFRHLQTLKSHVRIHTGKPYHCDPGLHFNKSQLRLHLQKKG 503  
DB 595 LHTGERPFCALCGKSPFKRKNLKHQRIHTGERPYTCGECGSKSPYKESLKDHLR----- 650  
QY 504 AATNTKVHYHLGGP 518  
DB 651 -----VH---SGGP 656

RESULT 14  
US-09-362-123A-4  
Sequence 4, Application US/09362123A  
Patent No. 6451558  
GENERAL INFORMATION:  
APPLICANT: Cooke, Michael Paul

APPLICANT: Holness, Claire  
APPLICANT: Sirenk, Oksana  
TITLE OF INVENTION: No. 6451558el Genes in the Control of Hematopoiesis  
FILE REFERENCE: 4-30629A/SYS  
CURRENT FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: 09/128,310  
PRIOR FILING DATE: 1998-08-03  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 4  
LENGTH: 543  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-362-123A-4

Query Match 13.0%; Score 362; DB 4; Length 543;  
Best Local Similarity 27.1%; Pred. No. 4.9e-20;  
Matches 127; Conservative 53; Mismatches 183; Indels 106; Gaps 18;  
QY 102 RLSPTAPAVLAATYLOMEHVVOACH-----RFOASYE--PLGISLRPLE----- 146  
DB 86 RAFPDTPPGMTTS-----VFPVAGCHSVKSLQROGASPSREKRPYGVSVIYWERLLG 140  
QY 147 -----AEPPTPAPPPSGPRSE-----GHDPPTESRSCGQPPS--PAS 186  
DB 141 SGSGQASVSLRITSPLRPEGVRLAEKTLTEHALGRQPRPERQKFCAGVPGRTFGSA 200  
QY 187 PDPKACN-----WKTKYIYLSQASQAGSL-----VGENS--SGQPCPOARLPSG 230  
DB 201 QDLBAAGRGHHRMGAVQVEPRRLGQEPFTWDELGALHAGERSFCRCASKYFVSS 260  
QY 231 DEASSSSSSSSSEEGPIPGQOSRLSPTAATVQFKGAPASTYLLTQAQDTSGSPS 290  
DB 261 DLTKLRLHTG-----EPPYECAG-----CGKAFSQTSHLTQHRITSGEP 302  
QY 291 ERARLPSESEFSCQCEAVAGCSSGLDSLVPGE-----DKPYKQLCRSSFRYKGNLAS 346  
DB 303 -----YACPVCGKAPRHS--SLVANHQRITHAESKFRSECGKAPSHSNIISQ 348  
QY 347 HRTVTGKPYHCSIGAPFNRPNALKTSRIHSGEKYKCTGSRFYQVRSQPPSGFQ 406  
DB 349 HRKIHAGRPYACACQGRFCRNSHLIHERHTGKPFVCLCAAF---SQGSSLP- 403  
QY 407 GKPARGVGQKGFCSQRODLKSPSOVAHLRAVLIHTGKPYPCPTGTRPHLQTL 466  
DB 404 -HQGVHTGKRPALPQCGRAE---SHSNLTQHQLHTGERPRCYDCGAFKAVL 458  
QY 467 KSHVRIHTGKPYHCDPGLHFNKSQL-----RLHLRQKGAATNTKYH 511  
DB 459 LSHRIHTGKPFVCTQCGRAFRRERPALFHHQRIHTGKTVARSASLH 507

RESULT 15  
US-08-820-170A-10  
Sequence 10, Application US/08820170A  
Patent No. 5831058  
GENERAL INFORMATION:  
APPLICANT: Isutomu, FUJIMARA  
APPLICANT: Takechi, WATANABE  
APPLICANT: Masato, HORIE  
APPLICANT: Toyomasa, KATAGIRI  
TITLE OF INVENTION: HUMAN GENE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Saas  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/820,170A  
FILING DATE:  
CLASSIFICATION: 536  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 711 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-820-170A-10

Query Match 12.8%; Score 357.5; DB 2; Length 711;  
Best Local Similarity 32.0%; Pred. No. 1,5e-19;  
Matches 90; Conservative 29; Mismatches 113; Indels 49; Gaps 9;

QY 264 VQKCG-APASTPYLLTQAQDTSGSPSEPARPLPGSEPPSCONCEAVAGSSGLD---S 319  
DB 448 VCIECGAFIOKAHLIVHQRSHTGEKP-----YQCHNCGKSPISKSLDIHHR 495  
QY 320 LVPGDEDEKPYKCOLCSSPFYKGNLASHRTVHTGEKPYHCSIGCARFNRPNLKTSHRTH 379  
DB 496 IHVG--EKPYECSDCGKTFQKSHLNIHQKHGTERHVCSECGKAFNOKSILSMHORIH 553  
QY 380 SGEKPYKCECTGSRFVQVRSQPPSGFGKPEARGVGQKGGFCSSORODLKSPS---QY 435  
DB 554 TGEKPYKCECGKAFPS-----KSQFK-EHQRIHTGEKPYVCTECGKAFNGRSNFIHQI 607  
QY 436 AHR-----AHVLIHTGEKPYPCPTGTRFRHLQTLKSHVRIHTG 475  
DB 608 THTREPRFVYKCGKAFVQKSELITHQRTMGEKPYECLDCGKFSKKPOLKVRIRIHTG 667  
QY 476 EKPYHCDPGGLHFRHKSQRLHLRQKGATNTKHYHILG 516  
DB 668 ERPYVCECGKAFNRSNENKH-QTHTRDXYKCSYSVKG 707

Search completed: July 14, 2003, 18:24:31  
Job time: 8.81542 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 12.144 Seconds  
(without alignments)  
4965.877 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789  
Sequence: 1 MGSPAPBEGALGVREFTTRH.....RQKGAATVTKVYHILGSP 518

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2789	100.0	518	9	US-09-815-379-4
2	1095.5	39.3	706	10	US-09-107-058-2
3	1095.5	39.3	706	10	US-09-761-117-2
4	469.5	16.8	711	9	US-10-034-934-90
5	457.5	16.4	641	9	US-10-164-359-17
6	428	15.3	516	9	US-09-974-298-118
7	426.5	15.3	516	9	US-09-764-891-4337
8	419.5	15.0	468	10	US-09-840-787-17
9	417	15.0	406	10	US-09-764-864-976
10	413	14.8	252	10	US-09-815-669A-12
11	412	14.8	468	10	US-09-864-761-37991
12	405.5	14.5	481	10	US-09-864-761-37651
13	399	14.3	241	10	US-09-764-864-935
14	397	14.2	278	10	US-09-925-300-1323
15	394.5	14.1	492	10	US-09-864-761-42897
16	390.5	14.0	551	10	US-09-864-761-37065
17	390.5	14.0	610	9	US-09-989-920-212
18	380	14.0	393	9	US-10-057-408-12
19	388	13.9	910	9	US-09-908-153B-40

20	388	13.9	922	9	US-09-908-153B-42	Sequence 42, Appl
21	386.5	13.9	547	10	US-09-764-864-1065	Sequence 1065, Ap
22	384.5	13.8	439	10	US-09-864-761-33476	Sequence 33476, A
23	384	13.8	924	9	US-10-007-280A-231	Sequence 231, App
24	383	13.7	184	9	US-09-765-555-38	Sequence 38, Appl
25	383	13.7	184	9	US-09-765-555-75	Sequence 75, Appl
26	383	13.7	186	9	US-09-908-153B-36	Sequence 36, Appl
27	383	13.7	525	10	US-09-864-761-33653	Sequence 33653, A
28	382.5	13.7	333	10	US-09-764-864-958	Sequence 958, App
29	382	13.7	184	9	US-09-765-555-40	Sequence 40, Appl
30	382	13.7	359	10	US-09-764-864-996	Sequence 996, App
31	381.5	13.7	433	9	US-10-171-311-232	Sequence 232, App
32	381.5	13.7	717	9	US-09-881-578-2	Sequence 2, Appl
33	380	13.6	180	9	US-10-106-698-5544	Sequence 5544, Ap
34	379	13.6	184	9	US-09-765-555-42	Sequence 42, Appl
35	378	13.6	184	9	US-09-765-555-39	Sequence 39, Appl
36	376.5	13.5	284	10	US-09-864-761-36268	Sequence 36268, A
37	376.5	13.5	573	9	US-10-157-031-36	Sequence 36, Appl
38	376	13.5	284	10	US-09-764-864-932	Sequence 893, App
39	376	13.5	361	10	US-09-764-864-893	Sequence 46977, A
40	375.5	13.5	338	10	US-09-864-761-46977	Sequence 1076, Ap
41	374.5	13.4	281	10	US-09-764-864-1076	Sequence 991, App
42	374	13.4	196	10	US-08-764-864-991	Sequence 36336, A
43	373.5	13.4	375	10	US-09-864-761-36336	Sequence 41, Appl
44	373	13.4	184	9	US-09-765-555-41	Sequence 43107, A
45	372.5	13.4	614	10	US-09-864-761-43107	

## ALIGNMENTS

RESULT 1  
US-09-815-379-4  
Sequence 4, Application US/09815379  
Publication No. US20030073613A1  
GENERAL INFORMATION:  
APPLICANT: RASTELLI, LUCA  
TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS  
FILE REFERENCE: 10716/35  
CURRENT APPLICATION NUMBER: US/09/815.379  
PRIOR APPLICATION NUMBER: 2001-03-22  
PRIOR FILING DATE: 2000-03-22  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 518  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-815-379-4

Query Match 100.0%; Score 2789; DB 9; Length 518;  
Best Local Similarity 100.0%; Pred. No. 6.3e-148;  
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGSPAPBEGALGVREFTTRHSDVYGNELRLRGLTDVTLVGGQPLRAKAVLIACS 60  
DB 1 MGSPAPBEGALGVREFTTRHSDVYGNELRLRGLTDVTLVGGQPLRAKAVLIACS 60  
QY 61 GFYFIFPGRAVGVDTSLPGCPARGPAPLLDPMYLSRLSLSPATYPAVLAATYLIOM 120  
DB 61 GFYFIFPGRAVGVDTSLPGCPARGPAPLLDPMYLSRLSLSPATYPAVLAATYLIOM 120  
QY 121 EHVQACRHFIOASYEPLGISLRPLEAPPTPTAPPGSPRRSGHPPTESRSCSGG 180  
DB 121 EHVQACRHFIOASYEPLGISLRPLEAPPTPTAPPGSPRRSGHPPTESRSCSGG 180  
QY 181 PPSAPSPPKAKNMKKRYTYLVNSQASQAGSLVGRSSGQPCQARLPSSGDEASSSSSS 240  
DB 181 PPSAPSPPKAKNMKKRYTYLVNSQASQAGSLVGRSSGQPCQARLPSSGDEASSSSSS 240

QY 241 SSSSSSEGPPIPGOSRLSPTAAVQFKCGAPASTPYLLTSOQADTSGSPSEBARPLPGSE 300  
DB 241 SSSSSSEGPPIPGOSRLSPTAAVQFKCGAPASTPYLLTSOQADTSGSPSEBARPLPGSE 300  
QY 301 PFSCQNEAAGSSGSLDVLPGDEDEPKYKCOLCRSSFRKGNLASHRTYHTGKPYHCS 360  
DB 301 PFSCQNEAAGSSGSLDVLPGDEDEPKYKCOLCRSSFRKGNLASHRTYHTGKPYHCS 360  
QY 361 ICARFRPNALKTHSHIHGSEKPYKCTGSRFVQVRSOPSPSGFOGKPARGGVGOKGF 420  
DB 361 ICARFRPNALKTHSHIHGSEKPYKCTGSRFVQVRSOPSPSGFOGKPARGGVGOKGF 420  
QY 421 CSSQRODLKSPSQVAHLRAHVLHTGKEKPYPCPTGTRFRHLQTLKSHVRIHTGKPYH 480  
DB 421 CSSQRODLKSPSQVAHLRAHVLHTGKEKPYPCPTGTRFRHLQTLKSHVRIHTGKPYH 480  
QY 481 CDPGGLHFRKSQLRLHROKGAATYTKVHYHILGSP 518  
DB 481 CDPGGLHFRKSQLRLHROKGAATYTKVHYHILGSP 518

## RESULT 2

US-09-107-058-2  
Sequence 2, Application US/09107058  
Patent No. US20010010922A1  
GENERAL INFORMATION:  
APPLICANT: Dalia-Favera, Riccardo  
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC  
TITLE OF INVENTION: LOCUS bcl-6  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,058  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0576/43771-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 706 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-107-058-2

Query Match 39.3%; Score 1095.5; DB 10; Length 706;  
Best Local Similarity 35.7%; Pred. No. 1.4e-53; Index 271; Gaps 18;  
Matches 264; Conservative 60; Mismatches 144;

QY 1 MGSPAAEGALGVYRETRHSDVGLNLTNRGLITDVTLLVGGQPLRAHRAVLIACS 60  
DB 1 MASPA--DSCI-----QTRHARVLTNLNLTNRSDILTDVIVVNSBQFPAHKTIVLMAR 54  
QY 61 GPFYSIFRGAGVGVVDVLSLPGSGEARGFAPLIDFMTSRLSPATAPAVLAATYLOM 120  
DB 61 GPFYSIFRGAGVGVVDVLSLPGSGEARGFAPLIDFMTSRLSPATAPAVLAATYLOM 120

DB 55 GLFYSIFTDQLKCNLSVINLDEINPEGFCILLDFMTSRLNLREGNIMAVNATAYLOM 114  
QY 121 BHVYQACHRFIOAS----- 134  
DB 115 BHVDTCKRFKFASEAMVSAIKPREEFLNSRMLMPDIMA YRGREVENNLPLRSABG 174  
QY 135 -----YEPGLGIS----- 141  
DB 175 CSBRAPAPSLYGLSTPPASYSMTSHLPVSLFSDPEFQVDMVMPANPPKERALPCDS 234  
QY 142 -----LRTL----- 145  
DB 235 ARPVPGEYAPPTLEVPVNVCHSNISPKETIDPEARSDMNVAVABGLKAAPASARNAPYF 294  
QY 146 -----EAPPT-----PPTAP-----PPGSRSEGHDPPTES-----RSC 177  
DB 295 PCDKASKERERSSSEDELHFEPPAPLRKGLVSPSPQSDCOFNPPTLACSSKNAC 354  
QY 178 ----SQGPPSPASPDEKACNMKKYKIVLS-----QASQAGSLVGERSSGQPCP 223  
DB 355 ILQSGSPPAKSPPTDPAKCSMKYKFTVLNSLNQNAKPGCPQOAEIGRLSPRAYTAPAC 414  
QY 224 QARLPBGDEASSSSSSSSSEEGPIPGOSRL-----SPTATVQF----- 266  
DB 415 QPPM-EPENLDQSPKLSASGEDSTP-QASRLNNIVNRSMTGSRSSSESHSPLVMP 472  
QY 267 ----KCG--APASTPYLLTSQ-----AODTSGSPSEBARPLPSEFPSCQNEAAGSSG 316  
DB 473 PKCTSCGSSQSPQAEKCLHTAGPTFAEMGEFQSEISDSBSCENGAFECNCCRSEBES 532  
QY 317 LD-SLVPGDEDEPKYKCOLCRSSFRYKGNLASHRTYHTGKPYHCSICGARFNRPNALKTH 375  
DB 533 LKRHTLQTHSDPKYCDRCQASFRYKGNLASHRTYHTGKPYHCSICGARFNRPNALKTH 592  
QY 376 SRIHGEKPYKCTGSRFVQVRSOPSPSGFOGKPARGGVGOKGFCSSRODLKSPSPGV 435  
DB 593 TIIHGEKPYKCTGCAEFV-----QV 614  
QY 436 AHLRAHVLHTGKEKPYPCPTGTRFRHLQTLKSHVRIHTGKPYHCDPGLHFRKSQLR 495  
DB 615 AHLRAHVLHTGKEKPYPCPTGTRFRHLQTLKSHVRIHTGKPYHCDPGLHFRKSQLR 674  
QY 496 LHLROKGAATYTKVHYH 514  
DB 675 LHLROKGAATYTKVQYRV 693

## RESULT 3

US-09-761-117-2  
Sequence 2, Application US/09761117  
Patent No. US20010012887A1  
GENERAL INFORMATION:  
APPLICANT: Dalia-Favera, Riccardo  
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS  
TITLE OF INVENTION: bcl-6  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/761,117  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US-X  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 TELE: 422523 COOP UI  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 706 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-761-117-2

Query Match 39.3%; Score 1095.5; DB 10; Length 706;  
 Best Local Similarity 35.7%; Pred. No. 1.4e-53;  
 Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

1 MGSPAPBEGALGVREPTSSDVLGNLNLRLGILTVTLVGGQPLRAKAVLIACS 60  
 1 MASPA--DSCI-----QFTRHARDVLLNLRSLRDLITDVIVVSRQFRAKTVLMAWR 54  
 61 GEFYSIFGRAGVGVVLSLPGGPEARGAPLLDFMYTSKRLSPATAPAVLAATYLOM 120  
 55 GLPFYSIFTLQKCNLSVINLDPINPEGCILDFYTSFLNREGNIMAVATANYLOM 114  
 121 EHYVQACHRFIOAS----- 134  
 115 EHYVDCKRIRKASBEMVSAIKPRPEEFLNSRLMPDIDIMAYRGREVENNLPLRSADG 174  
 135 -----YEPFGIS----- 141  
 175 CSRAFAELYSGLSTPPASYSMTSHLPVSSLLFSDREEDVMPVAPNPFKERALPCDS 234  
 142 -----LRPL----- 145  
 235 ARPVEYSRPTLEVS PNVCHSNITSPKETIPEARSQMHYSVABELKRAAPARAPAYF 294  
 146 -----EAPPT-----PPTAP-----PPGSPRSEGHDPPTES---RSC 177  
 295 PCDKASKEEERSSSEDEIALHFEPPNAPLNKGLVSPQSKDCQPNPTEKSSGNAC 354  
 178 -----SGCPSPASPDKACNMKKYKTIYNS-----QASQAGSLVGERSSQPCP 223  
 355 ILQSGSPSPAKSPPTDPAKCSWKYKFIYVNSLNQNAKPGGPEQALGRLSPRAYTAPAC 414  
 224 QARLPFGDEASSSSSSSSSEEGIPGPQSRP-----SPTAATVQF----- 266  
 415 QPFW--EPENLUDLOSPTKLASGDSITP--QASRLNINIVNRSMTGSRSSSESHPLYMP 472  
 267 ----KCG--APASTPYLLTSQ-----AQTSGSPSEPARPLPGSEFPSCNCEAVACSSG 316  
 473 PXTSGSGSPGHAEWCLHTAGPTFAEBEGETQSEYSDSSCEGAFPCNCDRCDFSEAS 532  
 317 LD--SLVPGDEDPKYQOLCRSSFRYKGNLASHTVTGKRYKCSICGARPNRPAULKH 375  
 533 LKRNHTLQTHSDKRYKCDRCOASFRYKGNLASHTVTGKRYKCNICGAFNPAULKH 592  
 376 SRHSSEKPYKCEICGSRFVQVARSPPSGFGKPARAGVQKQKGFCSQRODLKSPSPQV 435  
 593 THHGEKPYKCEICGARFY----- 614  
 436 AHLRAVLIHTGKRYPCPTCGTRFRLHLOTLSKSHVRIHTGKRYKCDPCGLHFRHKSQR 495  
 615 AHLRAVLIHTGKRYPCBICGTRFRLHLOTLSKSHVRIHTGKRYKCDPCGLHFRHKSQR 674  
 496 LHLRQGAATNKYVYHI 514  
 675 LHLRQHGAIITNKVQIRV 693

RESULT 4  
 US-10-034-934-90

Sequence 90, Application US/10034934  
 Publication No. US20030068624A1  
 GENERAL INFORMATION:  
 APPLICANT: Recipon, Herve  
 APPLICANT: Sun, Yongming  
 APPLICANT: Chen, Sei-Yu  
 APPLICANT: Liu, Chenghua  
 APPLICANT: Turner, Leah  
 TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific  
 FILE REFERENCE: DEX-0245  
 CURRENT APPLICATION NUMBER: US/10/034,934  
 PRIOR FILING DATE: 2000-10-26  
 PRIOR FILING DATE: 2000-10-26  
 PRIOR FILING DATE: 2000-10-26  
 PRIOR FILING DATE: 2000-11-20  
 PRIOR FILING DATE: 2000-11-22  
 NUMBER OF SEQ ID NOS: 142  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 90  
 LENGTH: 711  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-034-934-90

Query Match 16.8%; Score 469.5; DB 9; Length 711;  
 Best Local Similarity 26.0%; Pred. No. 9.7e-19;  
 Matches 162; Conservative 67; Mismatches 220; Indels 173; Gaps 22;

23 DVLGNLNLRLGILTVTLVGGQPLR-----AKAVLIACGFFYSIFRGAGV-GUDV 77  
 16 NLHMHMLRLGLHGLCDVTAVSYEQVGRDFMAHRAVLAATSKPFKEVFLNKSVDGRT 75  
 78 LSLPGPEARGAPLLDFMYTSKRLSPATAPAVLAATYLOMENVVQACHRFIOASVYP 137  
 76 NYVLAENVQVADPASTLFEVYTAQVVEEDVQRMLEVAEKLCLDLSCTFOLKQMLBS 135  
 138 LGISLRPL-----EAPPT-----PPTAPPPGSPRSEGHDPPTES---RSCSGPSP 184  
 136 VALLEQNTSESGEVEVSSGSQVSAAPAPRASVATDGPSPGLTDLVDYGERBASNGMSD 195  
 185 ASPDPACNMKKYKTIY----- 213  
 196 LPPKSKDKLDKCKEVVPPYKIRASGRILGRKYLVEIPKKYTRSLREQQTAEGDV 255  
 214 GERSSGQCPQARLPSG----- 253  
 256 GD-----YRCPDQSPDRVGTMEQVSKNEGCOAGALEBELSKKAPBESSEEDDEGE 311  
 254 QSRLEPTATVQFC----- 289  
 312 KKKSN-----FKSICEKAFLEYKSPFLKSKRHHGVAATEVYVC-----DTCQTFAN 359  
 290 ----SERAPPLPGR--PFSQNC-----EAVACSSGGLSLVPGDEDPKYQOL 333  
 360 RCNLKSHQRYHVASSSRHPPELCCGKKFRKDKVKNHVLQVHEG-----GG--RHRQ 411  
 334 CRSSFRYKGNLASHTVTGKRYKCSICGARPNRPAULKHRSISGEKPYKCEICGSR 393  
 412 CGKGLSSKTALRLHRTHTHTGDRPYCTEGGARFSPSLDKTHMRHTGKRPVCECGAR 471  
 394 FVQVARSPPSGFGKPARAGVQKQKGFCS-----SORDLK-----SP----- 432  
 472 FTQ-----NMHLIYHKGCHTGERPFWCEITGCKSPASKEVLYLKHNRHTGSRPFCEVCF 525  
 433 ---SOVAHLRAVLIHTGKRYPCPTCGTRFRLHLOTLSKSHVRIHTGKRYKCDPCGLHFR 489  
 526 RTFAQRNLSYOHIKVHTGERPYCCDQCGKQFQTLNALORHRIHTGSRPFMCAAGRTFT 585

QY 490 HKSQRLHLRQKHGAATNTKVA 511  
Db 586 DKSTLRHH-----TSIH 597

## RESULT 5

US-10-164-359-17  
; Sequence 17, Application US/10164359  
; Publication No. US2003001276A1  
; GENERAL INFORMATION:  
; APPLICANT: Chin, Khew-Yoon  
; TITLE OF INVENTION: Nucleic Acid and Protein Expressed Thereby and Their Involvement  
; FILE REFERENCE: 601-1-108US  
; CURRENT APPLICATION NUMBER: US/10/164,359  
; PRIOR FILING DATE: 2002-08-06  
; PRIOR APPLICATION NUMBER: PCT/US00/33438  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/169,418  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-359-17

Query Match 16.4%; Score 457.5; DB 9; Length 641;  
Best Local Similarity 25.0%; Pred. No. 4e-18;  
Matches 168; Conservative 56; Mismatches 195; Indels 253; Gaps 25;

QY 7 PEGALGVREPTSHSDVNLNLELRLR-GILTDVTLVGGOPLRAHKAVALIAGSFPFY 65  
Db 11 PSCC--YTYVSHSTHSTMLNINQKNGRCFCDVLRVDBEPRAHRAALACSIFYES 68  
QY 66 IFRGRAGVG-----VDV---LSLPGCP-----EARGFAPLIDFMYTSLRL 103  
Db 69 VFSAQLGDGGADGADGADGATAPGCGAGSRELEMTISSKVFGDIDFAYTSRIYV 128  
QY 104 SPATRAVLAALAYTLQMEHVQACHRTQNSYEPGLISLPLAEPTPTAP-----P 157  
Db 129 RLBSFELMTAAFLMLRSVIEICQEVTKQS-----NVQILVPPARADIMLFR 177  
QY 158 PGSPRSEGHPPPTESRSC---SQGPPSPAPDPKAKMKKYIYVLSQASQASLVG 214  
Db 178 PGT--SDLGFPPLMTGCAALAAANSNGIAGSMQBEER-----AAAGAAIA 221  
QY 215 ERSSGQPCPGA-RLPSGDEASS--SSSSSSSSSEEGPIPGQSRISPTAATVQFCGA 270  
Db 222 GQASLPVLPFVDRLPMVAGPLSPOLLTSPPPSVASSAPLTKRGRGRPRKAN----- 274  
QY 271 PASTPVLTLSQADDTGSSPS--ERARLP-----GSEFSGQNC---EAVAGCS---GLD 318  
Db 275 -----LDSM---FGSPGLREAGILPCGLGKVFDTANRLROHQAQHVTSIQLGVI 324  
QY 319 SLVP-----GDEDEKPYK-----COLCRSPFYKGNLASHRTVVTGKPYH 358  
Db 325 DLPPRLGNGELPISBDPGPRKRSRTRKQVACETICGKIFRDVYHLNRKLSISGKRPIS 384  
QY 359 CSICGARFNRPANLKTSHRIHSGE--KPYKCTGSRFVQVRSQPSGFGCKPARAGVQ 416  
Db 385 CPVCGARFKRKDMYSYVRSHDSVGKPYICSGCGKF----- 422  
QY 417 KGFCSGQRODLKSPSQVNLAAHY-LITGKPYPCPCGTRFRLQTLKSHVRIH-- 473  
Db 423 -----SRPDLNGHKQVITSERPHKQPCYNASFATRLRSLHACHED 466  
QY 474 ----- 473  
Db 467 KVPQVCGKTLRAAYADHLKHSSEBPSNCSICNREGQCKSHQDPIESSDSYIGDLSAS 526

QY 474 -----TGKPYHCDPCGLHFRHKSQRLHLRQKHGAAT 506  
Db 527 DLKTEPKQANGSPSCDMVAPRKHMESDEKTKPCBEGSPFRSXSILNKHIQ----- 579  
QY 507 NTKVHYHLIGP 518  
Db 580 --KVHVRALGCP 589

## RESULT 6

US-09-974-298-118  
; Sequence 118, Application US/09974298  
; Patent No. US20020156263A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Huei-Mei  
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
; FILE REFERENCE: PA-0037 P  
; CURRENT APPLICATION NUMBER: US/09/974,298  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,331  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 118  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1  
US-09-974-298-118

Query Match 15.3%; Score 428; DB 9; Length 516;  
Best Local Similarity 34.6%; Pred. No. 1.4e-16;  
Matches 110; Conservative 35; Mismatches 111; Indels 62; Gaps 12;

QY 223 PQARLP-----SGDEASSSSSSSSSEEGPIPGQSRISPTAAT--VOFK 267  
Db 21 POSHLPELQFOAEKGIYKYDMEKSVNSS-----LVSPQRISSTVTHISHTYE 72  
QY 268 CGAPASTPVLTLSQADDTGSP-----SEARLP-PSSE-----PSCQNCERV 310  
Db 73 CNPVD--LFTQKEKANIETGTHYKNERGKAFHOGILHTTIOIHTKETQKDCIGKI 129  
QY 311 ACCSSGLDS---LVGDEDEKPYKCOLCRSPFYKGNLASHRTVHTGKPYHCSIGAREN 367  
Db 130 FKKSNLASHQRIHNG--EKPYKNECGKVFFHNSHLAQHRIHTGKPYKNEGKVEN 187  
QY 368 PRANLKTSHRIHSGKPYKCTGSRFVQVRSQPSGFGCKPARAGVGQKGFCSQROD 427  
Db 188 QISHLAQHRIHTGKPYKNECGKVFFHNI-----SHLAQHRIHTGKPYKNECGKV 241  
QY 428 LKSPSQVNLAAHYLHTGKPYPCPCGTRFRLQTLKSHVRIHTGKPYHCDPCGLH 487  
Db 242 F---SRNSYLVOHLIHTGKPYKNCVCGKVFHHISHLAQHRIHTGKPYKNECGKV 297  
QY 488 FRHKSQ-----RLHLRQK 501  
Db 298 FSHKSLVNMHRIHTGK 315

## RESULT 7

US-09-764-891-4337  
; Sequence 4337, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; OTHER INFORMATION: data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4337  
LENGTH: 127  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (119)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (120)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (127)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4337

Query Match 15.3%; Score 426.5; DB 9; Length 127;  
Best Local Similarity 81.7%; Pred. No. 4e-17;  
Matches 89; Conservative 5; Mismatches 12; Indels 3; Gaps 2;

QY 1 MESPAPBAGALGVREFTTHSSDVLGNLNLRLRGLTDTVTLVGGQPLRAKAVLIACS 60  
DB 16 MESPAPBAGALGVREFTTHSSDVLGNLNLRLRGLTDTVTLVGGQPLRAKAVLIACS 75  
QY 61 GFFYSIFRGRAGVGVVLSLPGGPARGAPLIDFMYTSLRLSPATAP 109  
DB 76 GFFYSIFRGRAGVGVVLSLPGGPARGAPLIDFMYTSLRLSPATAP 121

RESULT 8  
US-09-840-787-17  
Sequence 17, Application US/09840787  
Patent No. US20020058264A1  
GENERAL INFORMATION:  
APPLICANT: Lai, Preeti  
Hillman, Jennifer L.  
Bandman, Olga  
Shah, Purvi  
Au-Young, Janice  
Yue, Henry  
Guegler, Karl J.  
Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/840,787  
FILING DATE: 23-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/518,865  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT03  
CLONE: 641127  
SEQUENCE DESCRIPTION: SEQ ID NO: 17 :  
US-09-840-787-17

Query Match 15.0%; Score 419.5; DB 10; Length 488;  
Best Local Similarity 29.5%; Pred. No. 4e-16;  
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

QY 176 SCSQGPSPSPAPDPKACN-----KKKKYIYLNLSA-----SQAGSLVGRSS 218  
DB 137 TCQNG-QLKESLDPIDCKDIHGWK--SQVSCSQRGHTBEKPCDHNNCKIINTSPD 193  
QY 219 GQPCQARLPSCDEASSSSSSSSSEGGPIPGQSRSLPYATVQFKCAPASTPYLL 278  
DB 194 GHPYKIHAEKQYSGSCGNKFSQSS-----LL 223  
QY 279 TSQADTSGSPERARPLPGESEFSCQNC-EAVACSSGLDVLVGEDKPYKCOLCRSS 337  
DB 224 LHQDHTTEKP-----YKCEOCCKGFTRSSSLIHQAVHTDEKPYKCDKCGK 271  
QY 338 FPKYNLMSHRVHTGKPYHCSIGARFNRPANLKTSHRISGKPKCTGSRFVQ- 396  
DB 272 FTRSSSLIHAVVHTGKPYKCDKCGKFSQSKLHQRVHTGKPYCECEGKNSFSQR 331  
QY 397 ----VRSQPPSGFGKPARGVGVGKGFCSQRODLKSPSQVAHLRAHVLITGKPY 452  
DB 332 SNLHIHQVHTG--BRPYKCGCKGF-----SSSNLHIHRCITHTGKPYQ 376  
QY 453 CPTGTRPRRHQLTKSHVRIHTGKPYHCDPCGLHFRHKSQI---RLHLRQK----- 501  
DB 377 CYCEGKGFSSQSDLIHRLVHTGKPYHCGKGFSSQSKLHQRVHTGKPYECSKC 436  
QY 502 -HGATNTKRVHY 513  
DB 437 GKGFSSQSNLHIH 449

RESULT 9  
US-09-764-864-976  
Sequence 976, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 976  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-864-976

Query Match 15.0%; Score 417; DB 10; Length 406;  
Best Local Similarity 36.6%; Pred. No. 4.5e-16;  
Matches 94; Conservative 26; Mismatches 105; Indels 32; Gaps 7;

QY 267 KCGAPAS-TPYLLTQADDTSGSPSE-----RAPPSPGSPFSCONCEA 309  
DB 77 ECGKSNMNSHLIWQXHTGKPYECPGCGAFSEKSRRLRHQGTHTGKPYKDCDCX 136  
QY 310 VAGCSSGL--DSLVPGBDKPYKCOLCRSSFRYGNLASHRTVHTGKPYHCSIGARF 366

Db 137 AFSAGSLRIHQTHTG--EKPECHCKGSPYKSTILVHQTHGKEPECEKCKSF 194  
 Qy 367 NRPAUKTHSRIRSGEYKCTGSRFVGRSPGSGFQKPARGVGQKGFSSQRQ 426  
 Db 195 SHHSGLRNHRRTHTGERPYKDECGKAF-----KLMSGLR-KHRTHTGKPYKQCKG 248  
 Qy 427 DLKSPSQVAHLRAVLIHTGKPYPCPTGTRPRHLQTLKSHVRHTGKPYKPCGL 486  
 Db 249 AF-----GQKSLGRHHRHTGKPYKCNHCBKAFSGKSNLRVHHRHTHTGKPYQCECGK 304  
 Qy 487 HERNKGLRLRLKQKHG 503  
 Db 305 TFRKSNLRGHRHTHTG 321

RESULT 10  
 US-09-816-669A-12  
 ; Sequence 12, Application US/09816669A  
 ; Patent No. US20020137019A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARABEDIAN, Michael  
 ; APPLICANT: TANEJA, Samir  
 ; APPLICANT: HITTELMAN, Adam  
 ; APPLICANT: MARKIS, Steven  
 ; TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF  
 ; TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL CO  
 ; FILE REFERENCE: GARABEDIAN=1.1A  
 ; CURRENT APPLICATION NUMBER: US/09/816,669A  
 ; CURRENT FILING DATE: 2001-03-26  
 ; PRIOR APPLICATION NUMBER: 60/225,618  
 ; PRIOR FILING DATE: 2000-08-15  
 ; PRIOR APPLICATION NUMBER: 60/191,768  
 ; PRIOR FILING DATE: 2000-03-24  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 12  
 ; LENGTH: 252  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-816-669A-12

Query Match 14.8%; Score 413; DB 10; Length 252;  
 Best Local Similarity 37.2%; Pred. No. 4,6e-16;  
 Matches 99; Conservative 25; Mismatches 96; Indels 46; Gaps 8;  
 Qy 265 QPKGAPASTPYLLTSGAQTSGSPSRARPLPSEFSGNCEBAVAGSSGL---DSL 321  
 Db 1 EFQCG---KVFQATSQL-----AHMVRHTGKPYKNDGGRAPSDRSILTFHQAH 49  
 Qy 322 PGDEDRPYKQOLCRSSFRYKGNLASHRTVATGKPYKCSICGARFNRPAUKTHSRIRSG 381  
 Db 50 TG--EKRYKCHGEGKVFRRHNSYLATHRRHTGKPYKCNCGAFHSHSLTTHKVIHNS 107  
 Qy 382 EKRYKCTGSRVYVRSQPPSGFQKPARGVGQKGFSS-----SORDL----- 428  
 Db 108 EKRYKCNQCGKVFQ-----NSHLNHRHTHTGKPYKNECKKAFSVYSSILTTQAH 161  
 Qy 429 --KSP-----SOVAHLRAVLIHTGKPYPCPTGTRPRHLQTLKSHVRHTGK 477  
 Db 162 TGKKPYKCNCEGKVFQONAHNHRHTHTGKPYKCTCGKAFRRVNSLTTHTHHTGK 221  
 Qy 478 PYHCDPCGLHFRHKSQRLRLKQKHG 503  
 Db 222 RYKCNCKGKVFRRSSNLSHHRHTG 247

RESULT 11  
 US-09-864-761-37991  
 ; Sequence 37991, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 FILE REFERENCE: Aemica-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263,6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00669  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00668  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00663  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00662  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00661  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 37991  
 LENGTH: 468  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AC005261.1  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7  
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
 OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 0.00e+00  
 OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALUATE 1.00e-105  
 OTHER INFORMATION: EST\_HUMAN HIT: AF150239.1, EVALUATE 1.00e-105  
 US-09-864-761-37991

Query Match 14.8%; Score 412; DB 10; Length 468;  
 Best Local Similarity 29.3%; Pred. No. 9,9e-16;  
 Matches 114; Conservative 21; Mismatches 158; Indels 96; Gaps 9;  
 Qy 184 PASPPKACNMKKYKTYLVNSQASQAGSVGRSSGCPQRLRLSGDEASSSSSSSSS 243  
 Db 5 PRSTPTTCALISGISIFWGLTQAGS--GDSQLGQPKDQ---DGFSEWQGERLRPGL 58



[illegible]

RESULT 14  
US-09-925-300-1323  
Sequence 1323, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PAL01  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1323  
LENGTH: 278  
TYPE: prt  
ORGANISM: Homo sapiens  
US-09-925-300-1323

	14.2%;	Score 397;	DB 10;	Length 278;
Query Match				
Best Local Similarity	35.8%;	Pred. No. 3.9e-15;		
Matches	82;	Conservative 21;	Mismatches 78;	Indels 48; Gaps 3
QY	298	GSEFFSCQN-----CEAVAGCSSGLDLSVPDEDEKPYCCOLCRSFFRYKGNLASHRTVA	351	
Db	45	GSSSWDIQNSFREKLEQKSPSKTLEQDSPEVRORVYECQCKECSFPKGSLSLHERIH	104	
QY	352	TGKKPYNCSICGARFNRANLKTHTSRHSGEYPCYCEFCGSRFOVYNSQPPSGGQPAR	411	
Db	105	TGGPPECTHCGSFPAKKNLVTHORIHGEKPYQCKECSGSF-----	147	
QY	412	GGVGQKGGFCSSGQRDLKSPSQVLAHLNVLHNGEYRYPCEPTCGTRFRHLQTLKSHVR	471	
Db	148	-----SQRSLVAHEELHTGQKPYECALCQSPFRNQSILAVHRR	186	
QY	472	IHNGEKPYHNDPCGLHFRHKSQLRLHLKQKG---AATNTKATYHNLG	516	
Db	187	VHSGEKPYRCDQCGKAFSQKGSGLVHINRHTGLKPYACTQCRKSPHRRG	235	

```

RESULT 15
US-09-864-761-42897
; Sequence 42897, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
;
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;
; FILE REFERENCE: Aeomica-X-1
;
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIORITY APPLICATION NUMBER: US 60/180,312
; PRIORITY FILING DATE: 2000-02-04
;

```

```

1 PRIOR APPLICATION NUMBER: US 60/207,456
2 PRIOR FILING DATE: 2000-05-26
3 PRIOR APPLICATION NUMBER: US 09/632,366
4 PRIOR FILING DATE: 2000-08-03
5 PRIOR APPLICATION NUMBER: GB 24263.6
6 PRIOR FILING DATE: 2000-10-04
7 PRIOR APPLICATION NUMBER: US 60/236,359
8 PRIOR FILING DATE: 2000-09-27
9 PRIOR APPLICATION NUMBER: PCT/US01/00666
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00667
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00664
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00669
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00665
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00668
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00663
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00662
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: PCT/US01/00661
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00660
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: US 60/234,687
30 PRIOR FILING DATE: 2000-09-21
31 PRIOR APPLICATION NUMBER: US 09/608,408
32 PRIOR FILING DATE: 2000-06-30
33 PRIOR APPLICATION NUMBER: US 09/774,203
34 PRIOR FILING DATE: 2001-01-29
35 NUMBER OF SEQ ID NOS: 49117
36 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
37 SEQ ID NO 42897
38 LENGTH: 492
39 TYPE: PRT
40 ORGANISM: Homo sapiens
41 FEATURE:
42 OTHER INFORMATION: MAP TO ACC007228.1
43 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
44 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
45 OTHER INFORMATION: SWISSPROT HIT: P10078, EVALU = 0.00e+00
46 OTHER INFORMATION: EST_HUMAN HIT: AL062491.2, EVALU = 0.00e+00
47 OTHER INFORMATION: EST_HUMAN HIT: AL062491.2, EVALU = 0.00e+00
48 US-09-684-761-42897

```

	Query Match	14.1%	Score 394.5	DB 10	Length 492
	Best Local Similarity	37.3%	Pred. No. 9.8e-15		
	Matches 78	Conservative 23	Mismatches 65	Indels 43	Gaps 4
Qy	298 GSEFFSPCONCEAVAGSSGLDS--LYGDEDEKPKYKQCLRSFRRYKGNLASHRTVHTGE	354			
Db	201 GEKPFPCIDCGKAFSDHILGNQHRRIHTG--EKPKYKCDVCHKSPRYGSSLTVHQRIHTGE	258			
Qy	355 KPHVCSICGARFNRRPANI.KTISRTHSGEKPYPKCTCGSRFYQVRSQPPSPGCKPRAGCV	414			
Db	259 KPEPCVCKKATSHHASTLTHORVHSGEKPKCECGKAF-----	298			
Qy	415 GQKGFCSQRODLKSPPSOVAHLRAHNL.IHTGKPYPCETGTRFHHLTOLTKSHVR.IHT	474			
Db	299 -----RQNI-----HLASHRIHTGKPEPCACGKSFSSISQ.LATHOR.IHT	340			
Qy	475 GEKPYHCDPCGLHFRHKSQLRLHLRKQKG	503			
Db	341 GEKPYECKVCSAFTOKALHNOHQHTHTG	369			

Search completed: July 14, 2003, 18:26:57  
Job time : 14.144 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:07:56 ; Search time 24.0314 Seconds  
(without alignment)

4441.379 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789  
Sequence: 1 MGSPAPPGALGYREFTRH.....RQKQKATNTKVNHLGCP 518

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_TREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_rv1rus:\*  
17: sp\_bacterioplasmid:\*  
18: sp\_bacteriophage:\*  
19: sp\_bacteriophage:\*  
20: sp\_bacteriophage:\*  
21: sp\_bacteriophage:\*  
22: sp\_bacteriophage:\*  
23: sp\_bacteriophage:\*  
24: sp\_bacteriophage:\*  
25: sp\_bacteriophage:\*  
26: sp\_bacteriophage:\*  
27: sp\_bacteriophage:\*  
28: sp\_bacteriophage:\*  
29: sp\_bacteriophage:\*  
30: sp\_bacteriophage:\*  
31: sp\_bacteriophage:\*  
32: sp\_bacteriophage:\*  
33: sp\_bacteriophage:\*  
34: sp\_bacteriophage:\*  
35: sp\_bacteriophage:\*  
36: sp\_bacteriophage:\*  
37: sp\_bacteriophage:\*  
38: sp\_bacteriophage:\*  
39: sp\_bacteriophage:\*  
40: sp\_bacteriophage:\*  
41: sp\_bacteriophage:\*  
42: sp\_bacteriophage:\*  
43: sp\_bacteriophage:\*  
44: sp\_bacteriophage:\*  
45: sp\_bacteriophage:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2272	81.5	474	11	088282
2	509.5	18.3	539	4	015156
3	509.5	18.3	539	4	096EP2
4	482	17.3	610	4	09NCP7
5	478	17.1	610	11	099MD8
6	472.5	16.9	582	11	09CKJ8
7	469.5	16.8	711	4	09H116
8	457.5	16.4	641	11	09JMG9
9	457.5	16.4	641	11	09JMG9
10	447	16.0	687	4	09HBE1
11	442.5	15.9	703	11	09R161
12	442	15.8	537	4	09HBE2
13	441	15.8	470	11	09PVP8
14	439.5	15.8	569	11	09QZ48
15	439.5	15.8	545	4	09H117
16	437.5	15.7	565	11	088939

17	436.5	15.7	725	4	096S24	096S24 homo sapien
18	432.5	15.5	555	11	062518	062518 mus musculu
19	430.5	15.4	555	11	0923D1	0923D1 mus musculu
20	428.5	15.4	616	4	09Y529	09Y529 homo sapien
21	428.5	15.4	673	4	08TAL4	08TAL4 homo sapien
22	426.5	15.3	499	4	096CL5	096CL5 homo sapien
23	426.5	15.3	499	4	096CL5	096CL5 mus musculu
24	425	15.2	488	4	096R00	096R00 mus musculu
25	423	15.2	488	4	096R00	096R00 mus musculu
26	423	15.2	488	4	096R00	096R00 mus musculu
27	422.5	15.1	395	4	091W94	091W94 mus musculu
28	422	15.1	395	4	091W94	091W94 mus musculu
29	420.5	15.1	673	11	09QVT4	09QVT4 mus sp. mpl
30	420	15.1	607	11	09D4A7	09D4A7 mus musculu
31	419.5	15.0	637	4	096IR2	096IR2 mus musculu
32	419	15.0	516	4	096SV5	096SV5 mus musculu
33	419	15.0	531	11	09D5K8	09D5K8 mus musculu
34	418.5	15.0	488	4	08TAS5	08TAS5 mus musculu
35	418.5	15.0	488	4	095365	095365 mus musculu
36	418	15.0	665	4	08TDL0	08TDL0 mus musculu
37	417.5	15.0	569	4	099K61	099K61 mus musculu
38	417.5	15.0	569	4	099K61	099K61 mus musculu
39	417.5	15.0	701	4	09UG02	09UG02 mus musculu
40	417.5	15.0	701	4	09UG02	09UG02 mus musculu
41	416.5	14.9	744	11	062788	062788 mus musculu
42	415.5	14.9	480	11	061117	061117 mus musculu
43	415.5	14.9	686	4	09NR94	09NR94 mus musculu
44	415.5	14.9	803	4	096TB6	096TB6 mus musculu
45	415.5	14.9	874	11	09R163	09R163 mus musculu

## ALIGNMENTS

RESULT 1  
ID 088282 PRELIMINARY: PRT: 474 AA.  
AC 088282;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE BAZF.  
GN BAZF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98298267; PubMed=9632807;  
RA Okabe S., Fukuda T., Ishibashi K., Kojima S., Okada S., Hatano M.,  
RA Ebata M., Saisho H., Tokuhisa T.;  
RT "BAZF, a novel bcl6 homolog, functions as a transcriptional  
RT repressor";  
RT Mol. Cell. Biol. 18:4235-4244(1998).  
CC 1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL: AB011665; BAA31223.1; .  
DR HSSP: P15822; 1B80.  
DR MCD: MGI:1278332; Bazf.  
DR InterPro: IPR000210; BTB\_POG.  
DR InterPro: IPR000822; Znf\_C2H2.  
DR Pfam: PF00651; BTB\_1.  
DR Pfam: PF00096; Zf\_C2H2; 5.  
DR PRINTS: PRO0048; ZINC\_FINGER.  
DR SMART: SM00225; BTB; 1.  
DR SMART: SM00355; ZNF\_C2H2; 5.  
DR PROSITE: PS00977; BTB\_1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_5.  
DR PROSITE: PS01577; ZINC\_FINGER\_C2H2\_2; 5.  
KW DNA-binding; Nuclear protein; Zinc-finger.  
SQ SEQUENCE 474 AA; 51339 MW; 94886AB265B018D8 CRC64;  
Query Match 81.5%; Score 2272; DB 11; Length 474;

Best Local Similarity 83.4%; Pred. No. 5,9e-161;  
Matches 432; Conservative 11; Mismatches 31; Indels 44; Gaps 3;

```

QY 1 MGSPAPGALGYVREFTRHSDVLTGNINELRLRGITLDTVLTVGGQPLRAHKVLIACS 60
DB 1 MGSTAPGALGYVREFTRHSDVLTGNINELRLRGITLDTVLTVGGQPLRAHKVLIACS 60
QY 61 GPFYSIFRGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLSLPATAVLAATYLYOM 120
DB 61 GPFYSIFRGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLSLPATAVLAATYLYOM 120
QY 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPPAPRPSRSEGHDPPTESRSCSG 180
DB 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPPAPRPSRSEGHDPPTESRSCSG 180
QY 181 SPSPASDPKACNMKKYKYLIVNSQASGLVGRSSGQPCPARLPSPGDEASSSSSS 240
DB 181 SPSPASDPKACNMKKYKYLIVNSQASGLVGRSSGQPCPARLPSPGDEASSSSSS 240
QY 241 SSSSESEGPITPGPSRLSPTAATVQFKGAPASTPYLLTSQADTSGSPERARPLPGSE 300
DB 241 SSSSESEGPITPGPSRLSPTAATVQFKGAPASTPYLLTSQADTSGSPERARPLPGSE 300
QY 301 PFSCONCEAVAGSSGDLVPGDEDPKPYCOLCRSSFRYKGNLASHRTVHTGKPYRCS 360
DB 295 PFSCONCEAVAGSSGDLVPGDEDPKPYCOLCRSFRYKGNLASHRTVHTGKPYRCS 360
QY 361 ICGARFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGQGGF 420
DB 355 ICGARFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGQGGF 420
QY 421 CSSGRDLPKSPQVLAHRAHVLHTGKPYPCPTGCTRRHLOTLSHVRHTGKPYH 480
DB 390 -----QVLAHRAHVLHTGKPYPCPTGCTRRHLOTLSHVRHTGKPYH 436
QY 481 CDPGCLHFRHKSQRLHLRQKGAATNTKYHYHILGSP 518
DB 437 CDPGCLHFRHKSQRLHLRQKGAATNTKYHYHILGSP 474

```

RESULT 2

ID 015156 PRELIMINARY; PRT; 539 AA.

AC 015156;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Kruppel-related zinc finger protein hckrox.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKIN.FIBROBLAST;  
RX MEDLINE=98036076; PubMed=9370309;  
RA Widom R.L., Cullis I., Lee J.Y., Korn J.H.;  
RT "Cloning and characterization of hckrox, a transcriptional regulator  
of extracellular matrix gene expression";  
RL Gene 198:407-420(1997).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; AF007833; AAC51847.1; -  
DR HSSP; P08047; 1SP2.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF00651; BTB\_1.  
DR Pfam; PF00096; zf-C2H2; 4.  
DR PRINTS; PR00048; ZINCINGER.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 4.  
DR PROSITE; PS00097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_3.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 4.

KM DNA-binding; Nuclear protein; Zinc-finger  
SQ SEQUENCE 539 AA; 58218 MW; AE7C5D83C04D583F CRC64;

Query Match 18.3%; Score 509.5; DB 4; Length 539;  
Best Local Similarity 28.3%; Pred. No. 6.7e-30;  
Matches 160; Conservative 53; Mismatches 190; Indels 163; Gaps 18;

```

QY 1 MGSPAPGALGYVREFTRHSDVLTGNINELRLRGITLDTVLTVGGQPLRAHKVLIACS 60
DB 1 MGSP--EDDLIGI--PDPHSSBLSCNEGRQLGHLCDLTTRTGQLEVTRRAVLAACS 56
QY 61 GPFYSIF-----RGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLSL 103
DB 57 HYKKLFTGGGGAVVWAGSGTATGAGAGVCELDV-GPRLAG--ALLEAYATLTLT 113
QY 104 SPTAPAVLAATYLOMEHVVOACHRFIOASYEPLGISLRPLEAPPTP----- 152
DB 114 SSANPFAVLQAARLLEIPCVIAACWEIIQGS-----GL-----EAPSPDEDCERARQY 162
QY 153 ----PAPRPGSPRSEGHDPPTESRSCSGPPSPAPDPKACNMKKYKYLIVNSQASQ 208
DB 163 LEAFATATASGVPNGEDSPQVLP-----PPPPPPRPAVARSRRKRAFLDTKAR 215
QY 209 ASGLVGRSSGQPCPARLPSPGDEASSSSSSSSSEGPITPGPSRLSPTAATVQFKC 268
DB 216 ANHLVDE----VPTVAHPLTYEEBEVAGRVSSGGS-----GGDSYSPTGTASPE 265
QY 269 GAPASTPYLLTSQAD-----TSGSPERARPLPGSEPFSCQNC---EAVAGCS 315
DB 266 GPQSYEPYGEHEEELVYPAYGLAOGGP-----PLSPBELGSDDEDAIDPDLAAYLS 320
QY 316 -GLDVLVPG--DEDK-----PYCOLCRSSFRYKGNLASHRTVHTGKPYHCSIG 363
DB 321 LHQDVLAGDLSQDLYVKRSSQMPQECPCVKHIIHGCKLPRHRTHTGKPPACEVCG 380
QY 364 AEFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGQGGFCS 423
DB 381 VFTENDKLIKIMR----- 394
QY 424 QRQDLKSPQVLAHRAHVLHTGKPYPCPTGCTRRHLOTLSHVRHTGKPYHCDP 483
DB 395 -----KHTGERPSPCHPCARFLHSYDLKNHMLHTGDRPYECHL 434
QY 484 CGLHFRHKSQRLHLRQKGAATNTK 509
DB 435 CHRAFAKEDHLQRLHKQNCLEVRTR 460

```

RESULT 3

ID 096EP2 PRELIMINARY; PRT; 539 AA.

AC 096EP2;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Similar to kruppel-related zinc finger protein hckrox.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RA Straubeberg R.;  
RL Submitted (JUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC012070; AAH12070.1; -  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 4.  
DR PROSITE; PS00097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_3.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 4.

KW DNA-binding; Zinc-finger.  
SQ SEQUENCE 539 AA; 58026 MW; 8683616504B05B7E CRC64;

Query Match 18.3%; Score 509.5; DB 4; Length 539;  
Best Local Similarity 28.3%; Pred. No. 6.7e-30; Indels 163; Gaps 18;

Matches 160; Conservative 53; Mismatch 190; Indels 163; Gaps 18;

QY 1 MGSPAPAGALGVREFTTHSSDVLGNLNLRLGLITVTLVGGQPLRAHKAVALIACS 60  
DB 1 MGSP--EDDLIGI--PFPDHSSELISCLNDRGLGHCGLITRTGQLETRTHRAVALAAS 56  
QY 61 GFFYSIF-----RGRAGVDVLSLPGSPKARPAFLIDEMYSRLRL 103  
DB 57 HYFKKLFTEGGGAVWAGSGSGTATGAGAGVCELDLV--GPEALG--ALLEFAYATLT 113  
QY 104 SPATAPVALAATYLOMEHVQACHFIDASVEPGISLRPLEAEPT----- 152  
DB 114 SSANMPAVQAARLEPCVIAQMEILQS--GL-----EAPSPDEDCERARQY 162  
QY 153 ----PTAPPGSPRRSEGHDPPTESRSCSQGPPSPASFPKACNMKKYKVIYVNSQSQ 208  
DB 163 LEAPATATAGVPPNGEDSPQVLP-----PPPPPPRPARARRKPKAKFLQTKGAR 215  
QY 209 AGSLVGRSSGQPCPOARLPDGDEASSSSSSSSSESGPLPGPSRLSPATAVQFKC 268  
DB 216 ANHLVPE---VPTVAHPLVTEEBEVAAGRVSSGGS-----GPGDSYSPGTGASPE 265  
QY 269 GAPASPVYLTGQAD-----TSGSPSRAPPLGSEFFSQNC---EAVGCGSS 315  
DB 266 GQOSTEYEEGEEBELVYPAYGLAQGGP---FLPEELGSEEDLIDPLMVLSS 320  
QY 316 -GLDSLVPG--DEDK-----PYKQLCRSSFPYKKNLASHRTVHGEKPYHSG 363  
DB 321 LHQDNAPRLDQDQKLVKRRSQMPQECVCHKIHGAKSLPRHMTHTGKEKPFACEVG 380  
QY 364 ARPNRPANIKTSHRISGKPKYKCTGSRFVQVRSQPPSGQKPARGVGQKGFCS 423  
DB 381 VFTTRNDKLKIHMR-----KHTGERPYSCPNCAPAFILSHYDLKNHMLHTGDRPYE 394  
QY 424 QNODLSPSPQVAHLRAHVLHNGEKPYCPTCGTRFRHLQTLKSHVRIHTEKPYHCDP 483  
DB 395 -----KHTGERPYSCPNCAPAFILSHYDLKNHMLHTGDRPYE 434  
QY 484 CGLHPRHKSQRLRLRQKHGAATNTK 509  
DB 435 CHKAPAKEDHLQRHLKQNCLEVRTR 460

RESULT 4  
G9NPC7 PRELIMINARY; PRT; 610 AA.  
AC G9NPC7; 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Zinc finger protein SBB121 (Myoneurin).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN RA SEQUENCE FROM N.A.  
RA Zhang W., Wan T., Yuan Z., Cao X.;  
RT "A ZINC finger protein SBB121 expressed in human dendritic cells";  
RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
RN RA SEQUENCE FROM N.A.  
RP MEDLINE=20334280; Pubmed=10873615;  
RX Allel P.M., Seddigi N., Goudou D., Cifuentes-Diaz C., Romero N.,  
RA Velasco E., Rieger F., Perin J.P.,  
RT "Myoneurin, a novel member of the BTB/POZ-zinc finger family highly  
RT Blochem. Biophys. Res. Commun. 273:385-391(2000).

RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMERYO;  
RC Isocal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,  
RA Wagaitsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,  
RA Niomiya K., Iwayanagi T.;  
RT "NEO human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF076249; AAF80160.1; -  
DR EMBL; AF148848; AAF7138.1; -  
DR EMBL; AK021646; BAB13862.1; -  
DR HSSP; P08048; 7ZNP.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 8.  
DR PROSITE; PS50097; BTB; 1.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 8.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 8.  
KW DNA-binding; Metal-binding; Zinc-finger.  
SQ SEQUENCE 610 AA; 66861 MW; 842B62DDCC81480 CRC64;

Query Match 17.3%; Score 482; DB 4; Length 610;  
Best Local Similarity 26.8%; Pred. No. 8.5e-28;  
Matches 149; Conservative 65; Mismatch 212; Indels 130; Gaps 16;

QY 16 EFTRSSDVLGNLNLRLGLITVTLVGGQPLRAHKAVALIACGFFYSIRGRA--CY 73  
DB 2 QSHHCEHLERLNQREAGFLCDDCTIVIGEFQFAHNVNLASFSEYFQATIRSTENNV 61  
QY 74 GVDVSLPGSPARFAPLIDEMYSRLRLSPATAPVALAATYLOMEHVQACH----- 128  
DB 62 FLDDQGV---KADFQGLLEITVYTGTLNLSMNKELHQAADYKAVEVATKCKIKMED 117  
QY 129 -----RFTQASYEPLGISLRP-----LEAPPTPTAPPPG 159  
DB 118 FAFIANPSTETISITGNILNQTCCLTLRDYNNRSESVSTDLIQANPKGALAKSS 177  
QY 160 -SPRSSEGHDPPTESRSCSQP-----PSPASPPKACNMK---KY 197  
DB 178 QTKKKKKAFFNSPKTQONTVOYPSDILENASVELFLDANKLPTPVVEOVAQINDSELEL 237  
QY 198 KYIVNSQASQ--AGSLVGRSSGQPCPOARLPs---GDEASSSSSSSSSSSESGPIRG 252  
DB 238 TSVENTPPADIVITVYVKKRGSQPNCLAKHSMSNINASKPYEANGER--LDQ 295  
QY 253 PGRSLSPATAVQFKCAPASTPYLLTSGAODTSSPSEARAP---LGSEFFSQNCE 308  
DB 296 RYSKAKPMONT---CGKVFSEASSLRHMRHKG---VPPYCHLCGKAFTQCNQLK 346  
QY 309 AVAGSSGLDSLVPDEDEKPYKQLCRSSFRKKNLASHRTVHTG--EKPYHSGICGARN 367  
DB 347 THVRHTG-----EKPYKCELCDDGFAKCOLVFRSHRHHGEKPYKDVNCLQPA 397  
QY 368 RPANIKTSHRISGKPKYKCTGSRFVQVRSQPPSGQKPARGVGQKGFCSQORD 427  
DB 398 TSSNLKIRAKRKSQKPYVCDRCGRF----- 424  
QY 428 LKSPSPQVAHLRAHVLHNGEKPYCPTCGTRFRHLQTLKSHVRIHTEKPYHCDP 487  
DB 425 -----AQASTLYVHRHTGKEKPYVCDTCGKAFAVSSSLITSHRHTGKEKPYICGICGS 479  
QY 488 FHKSQLHLRQKHG 503  
DB 480 FISSGELNKHFRSHTG 495

```

RESULT 5
099MD8 PRELIMINARY; PRT; 610 AA.
ID 099MD8;
AC 099MD8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Myonaurin.
GN MYN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Alliel P.M., Seddigi N., Gondou D., Clifuentes-Diaz C., Romero N.,
RA Velasco E., Rieger F., Parin J.P.;
RA Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RT EMBL: AF349561; AK18605.1; -.
RL HSSP: P08048; 7ZNF.
DR MGD: MG1:1931415; Myrn.
DR InterPro: IPR000210; ZNF_C2H2.
DR InterPro: IPR000822; ZNF_C2H2.
DR Pfam: PF00051; BTB; 1.
DR Pfam: PF00096; ZF_C2H2; 8.
DR ProDom: PD000003; ZNF_C2H2; 1.
DR SMART: SM00225; BTB; 1.
DR SMART: SM00355; ZNF_C2H2; 8.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 610 AA; 68520 MW; 0C356BE1E00731F CRC64;

Query Match 17.4%; Score 478; DB 11; Length 610;
Best Local Similarity 26.8%; Pred. No. 1,7e-27;
Matches 150; Conservative 59; Mismatches 213; Indels 138; Gaps 14;

16 EFTRHSSDVGLNLELRGLTDLTVLVGGPPLRAHKAVALACSGFPYIFRGA--GV 73
2 QVSHCEHLERLNKQREAGFLCDCTVVGEPQKARNTVLAFFSYFGALVSTSENNV 61
74 GUDVLSLPGGPEKRGAPLIDENYTSRLRLSPATAVAALATYLOMEHYQAC----- 127
62 FLDSQGV---KADGFOKLLEFYITGLTLDNWNVEIHQADYLVVEEVTCKCKIMED 117
128 HRFIQASYEPLGLSLPRLAEPPPTAPPPG-----PRSEGHDPPTSRSC 177
118 FAFI-----ASPESTISITIGITELNQACCLTLTDYNNRRESEVSTDS 162
178 SQGPPSPAPDPACAKWKYKXIVLNSQAQASLGVERSGQPCFQARLPFGDEASSSS 237
163 VQANPKPRALTKKSSQSKKKKAFSSQKQKQ-----SKAVQYSDVLESASVLFLE 214
238 SSSSSSSSEEGPFPQOS-----RLSPPAATV-----QFKCG----- 269
215 TSKLSSPPVVEQIIQGNDSLELTIVVENTFPQDIQVTVAKRKRSQSHCALKEHSM 274
270 ---APASTPLYSQAODTSGSPSERARPL-----PGEFFSQ 305
275 SNIASVKSFELEENAGEILDAPS-KAKPWCNTCGVFSEASSLRHRMIRHKGVKPYVCH 333

```

```

306 NC-EAVAGCSSGLDLPDEDEKPYKCOLCRSSFRYKGNLASHRTVHTG-EXPHYCSICG 363
334 LCGKAFQCNQKHTVHTHTGERPYKCELCIDKGFQKQCLVHSHMHGEEKPYKCDVCN 393
364 AEPNRPANLKTSHRHSGEKPYKCEKCSRFVQVRSQPPSGQKGRPAKGVGQKGFCS 423
394 LQFAVTSNLIKIHARKHSGKPKVPCDRCGQRF----- 424
424 QKQDLKSPSQAHLRAVLIHTGEKPYKCPCTGFRHLQTLKSHVRLHTGEKPYKCDP 483
425 -----AQSSTLYHVRHTGKPKVCDTCGAFAVSSLLTHSRKHTGEKPYKCI 475
484 CGLHRRHKSQRLHURKXG 503
476 CCKSFISGELNKHFRSHTG 495

RESULT 6
099CXJ8 PRELIMINARY; PRT; 582 AA.
ID 099CXJ8;
AC 099CXJ8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE 281001IC24Rik protein.
GN MYNN OR 281001IC24Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Momberts P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlschki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK014315; BAB29266.1; -.
DR HSSP: P08048; 7ZNF.
DR MGD: MG1:1931415; Myrn.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000822; ZNF_C2H2.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF00096; ZF_C2H2; 7.
DR ProDom: PD000003; ZNF_C2H2; 1.
DR SMART: SM00225; BTB; 1.
DR SMART: SM00355; ZNF_C2H2; 7.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 582 AA; 65482 MW; 7415F8370CDDCB6F8 CRC64;

Query Match 16.9%; Score 472.5; DB 11; Length 582;
Best Local Similarity 26.4%; Pred. No. 4.1e-27;
Matches 150; Conservative 63; Mismatches 208; Indels 147; Gaps 15;

```

```

QY 16 EFTRSSDVLGNLNLRLGILTDVTLVVGOPRLAHKAVLACSGFFYSIRGRA--GV 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2 QVSHCHEHLERLNKREAGPLDCVIVGEGFKAHRYNLASFYSFGAIIYSTENN 61
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 74 GVDVLSLPGPEAKGAPLLDFMYTSLRLSPATAVLAATYLOMEHYVQAC----- 127
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 62 PLDSQV---KADGQKLEFLYGTGLMDSNVKEIHCADYLVVEEYVTKCKIMED 117
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 128 HRFIOGYPEPLGISLRLSEAPPTPTAPPPG-----PRSEGHDPPTSRSC 177
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 118 PAFI-----ASPESTISITGIELNQO?CLTLTDYNNREKSEVSTGS 162
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 178 SCGPSPAPDPKACMKKYKYLVLNSQASQAGSLVGRSSGQPCQALPSCDEASSS 237
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 VQANPQPALTLTKSSGSKKKKKGFSQKQGQ-----SKAVQYSDVLESVALLPLE 214
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 SSSSSSSSEGP1PGPQS-----RLSPATAV-----QFKCG----- 269
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 215 TSKLSPPVVEQIIOGNDSESELSTSVVENTFPQDIQYTVARKRKKSQSHCALKEHSM 274
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 270 ---APASTPYLLTSQADTSGSPSERAPL-----PSEFFSQ 305
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 275 SNIAVSKSPYLEENAGEELDAPPS-PAKPCNTCGVFSFASLRHMR1HKGVKPYVGH 333
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 306 NC-EAVAGCGSGLDLSPGDEDEKPYCOLCRSSFRYKGNLASHRTVHTG-EKPYHCSIG 363
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 334 LCGKATQCNQCKTHVTRHTGERPYKCELDCKGFAKQCVLFRSHNHGEKPYXCDVCN 393
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 364 AAFNRPAHLKTSRIHSGEKPYCETCGSRFVQVRSQPRGFGKPARGVGQKGFCS 423
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 394 LQFATSNLKIHKRKSSEKPYVCDTCGKAFVASSLITSRKHTGERP1CEL 424
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 424 QNDLSPSPQVAHLRAHVL1HTEKPYPCPTGTRFHLQTLKSHVRIHTEKPYHCDP 483
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 425 -----AQASTLYVHVRHNGEKPYCDTCGKAFVASSLITSRKHTGERP1CEL 475
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 484 CGLHFRHKSQRLRLHROKGAATNTKYH 511
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 476 CGNSTYDINLKKH-----KTKVH 494
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Query Match 16.8%; Score 469.5; DB 4; Length 711;
Best Local Similarity 26.0%; Pred. No. 8,6e-27;
Matches 162; Conservative 67; Mismatches 220; Indels 173; Gaps 22;

QY 23 DVLGNLNLRLGILTDVTLVVGOPRL---AHKAVLACSGFFYSIRGAGV-GVDV 77
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 16 NLHMEHLEIRLGHLCVTVSVVEYQGVKDFMAHRAVLAATSKPFEVPLNKSVDGTRT 75
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 78 LSLPGPEAKGAPLLDFMYTSLRLSPATAVLAATYLOMEHYVQACHRFIOASVSP 137
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76 NYVLEVQVADPASPFLFVYTAQVVEEDRVQRMLEVLEKLCLDSETCFOLKQMLSS 135
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 138 LGISIRPL-----EAPPT---PPTAPPPSPRSEGH-HPDPPTES-----RSCSQPSP 184
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 136 VLEELQNPESQEVLEVSSGSQVSAAPAPASVATGPHPSGLTDLDPYGERASNGMSSD 195
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 185 ASPDPKACMKKYKYL-----LNSQASQAGSLV 213
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 196 LPPKSKDLDKKEVVPYPKIRRASGLRAGKVFVFI PKKYTFRLRBOQKTAEGDV 255
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 214 GERSSGQPCQARLPSG-----DEASSSSSSSSSEEGPIPP 253
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 GD---YRCPOQSDPRVGTENEQVSKQEGAGLEBELSKAGPEEBEEDDEBEG 311
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 254 QSLSPATAVQFKC-----GAPASTPYLLTSQADTSGSP--- 289
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 312 KKKSN-----FKSICGKAPLYEKSPLKSKHGHVATEVYVC-----DTCCGTFFAN 359
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 290 ---SERAPFLPSG-PSQCNQ-----BAVAGCGSGLDLSPGDEDEKPYCOL 333
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 360 RCNLKSHQHNVASSRHHFPCBLCGKKPKKDKVGHV1QVHGG-----GGE--RHRQ 411
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 334 CASSRYRYKNLASHRTVHTGERPYHCSICGAFNRPAHLKTSRIHSGEKPYCETCGSR 393
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 412 CCKGLSSKTAALHRTHTGDRPYCTECGAPFQSPSLKTMRIHTEKPYVCECGAR 471
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 394 FVQVRSQPRGFGKPARGVGQKGFCS-----SORDLK-----SP----- 432
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 FTQ-----NHMLIYHKRCHTERPFCETCGKASAKYGLHNHR1HIGSRPKCEVCF 525
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 433 ---SQVAHLRAHVL1HTEKPYPCPTGTRFHLQTLKSHVRIHTEKPYHCDPGLHR 489
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 526 RTEAQNSLYGH1KHTGERPYCDQCGKQFLNALQHRHRIHTEKPYFCMAGCGRFT 585
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 490 HKSQRLHROKGAATNTKYH 511
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 586 DKSTLRRH-----TSIH 597
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 8
Q9UDU0
AC Q9UDU0 PRELIMINARY; PRT; 641 AA.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE WUGSC:H.DV40WZ3.1 protein (Zinc finger sarcoma gene long A isoform)
DE (BTR-POZ domain zinc finger transcription factor).
GN ZSG OR R1A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1107-1108(1998).
KW [2]
RP SEQUENCE FROM N.A.
RA Gattung S., Geisel C., Murray J.;

```

RT "The sequence of Homo sapiens PAC clone RP3-400N23.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.H.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20404868; Pubmed=10949935;  
 RA Mestazanglo T., Modena P., Tornelli S., Bullrich F., Testi A.,  
 RA Mezzalana A., Radice P., Azarelli A., Piloti S., Croce C.,  
 RA Pierotti M., Sozzi G.;  
 RT "A novel zinc finger gene is fused to EMS in small round cell tumor.";  
 RL Oncogene 19:3799-3804(2000).  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Chin K.-V., Yang W.-L., Kudoh K.;  
 RT "Novel CAMP Signalling via the Regulatory Subunit of the CAMP-  
 RT dependent Protein Kinase.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005003; AF01349.1; -  
 DR EMBL; AF254083; AAG09032.1; -  
 DR EMBL; AY028384; AAK19024.1; -  
 DR HSSP; Q05515; ICS3.  
 DR TRANSFAC; T04797; -  
 DR InterPro; IPR000637; AT\_hook.  
 DR InterPro; IPR000515; BPD\_transp.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR000822; ZnF\_C2H2.  
 DR Pfam; PF02178; AT\_hook.1.  
 DR Pfam; PF00651; BTB.1.  
 DR Pfam; PF00096; Zf-C2H2; 7.  
 DR Prodom; PD000003; ZnF\_C2H2; 1.  
 DR SMART; SM00384; AT\_hook.1.  
 DR SMART; SM00225; BTB.1.  
 DR SMART; SM00355; ZnF\_C2H2; 7.  
 DR PROSITE; PS00402; BPD\_TRANSFAC\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE; PS00097; BTB.1.  
 DR PROSITE; PS00354; HMGI\_Y.1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 6.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
 DR DNA-binding; Metal-binding; Zinc-finger;  
 KW SEQUENCE 641 AA; 69079 MW; 2951535380FOCE CRC64;  
 SQ  
 Query Match 16.4%; Score 457.5; DB 4; Length 641;  
 Best Local Similarity 25.0%; Pred. No. 6e-26;  
 Matches 168; Conservative 56; Mismatches 195; Indels 253; Gaps 25;

DB 275 -----LIDSM-----FGSGGLREAGILPCGLGKVFDTANRLRQHEAQHGYTSLQLYI 324  
 QY 319 SLVP-----GDEDKPKY-----COLCRSSFRRKGNLASRTVHTGKPRH 358  
 DB 325 DLPPRLGNGULPISDDPGPRKRSRTRKQVACEICGIFRVYHLNRLKLSHGSKPRS 384  
 QY 359 CSICGARFNRPANLTKTSHISGE--KPKCECTGSRFVQVNSOPPSGFGKPARGVQ 416  
 DB 385 CVCGGRFGRKORMSHVSHDSGVKPYICSGCGKF----- 422  
 QY 417 KSGFCSSQRDLKSPPSQVAHLPAHV-LIHTEKPYPCPTGTRPHLTLKSHVRIH-- 473  
 DB 423 -----SRPDHLNGHIKVYHTSERPHKCTCNASFATRDRLSLHACHED 466  
 QY 474 ----- 473  
 DB 467 KVPQVCGKYLRPAVADHLKHSRSPNSFCSIKNREGKCSHODPIESSDSYGLSDAS 526  
 QY 474 -----TSEKPYHCDPCGLAFPRHKSQTLRLRLRKHGAAT 506  
 DB 527 DKTPEKQANGSFQCDMAVPKMKHESDGEKKYPCPECGSFRRSKSYLNKHIIQ----- 579  
 QY 507 NTKVHYHILGSP 518  
 DB 580 --KVHVRALGSP 589  
 RESULT 9  
 Q9JMG9 PRELIMINARY; PRT; 641 AA.  
 ID Q9JMG9  
 AC Q9JMG9;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Transcription factor MAZR.  
 GN ZFP278 OR MAZR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RX MEDLINE=20136053; Pubmed=10669750;  
 RA Kobayashi A., Yamaguchi H., Hoshino H., Muto A., Sato K., Morita M.,  
 RA Hayaishi N., Yamamoto M., Igarashi K.;  
 RT "A combinatorial code for gene expression generated by transcription  
 RT factor bach2 and MAZR (MAZ-related factor) through the BTB/POZ  
 RT domain.";  
 RL Mol. Cell. Biol. 20:1733-1746(2000).  
 DR EMBL; AB029397; BAA90874.1; -  
 DR HSSP; Q05516; ICS3.  
 DR TRANSFAC; T04796; -  
 DR MGD; MGI:1891832; Zfp278.  
 DR InterPro; IPR000637; AT\_hook.  
 DR InterPro; IPR000515; BPD\_transp.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR000822; ZnF\_C2H2.  
 DR Pfam; PF02178; AT\_hook.1.  
 DR Pfam; PF00651; BTB.1.  
 DR Pfam; PF00096; Zf-C2H2; 7.  
 DR Prodom; PD000003; ZnF\_C2H2; 1.  
 DR SMART; SM00384; AT\_hook.1.  
 DR SMART; SM00225; BTB.1.  
 DR SMART; SM00355; ZnF\_C2H2; 7.  
 DR PROSITE; PS00402; BPD\_TRANSFAC\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE; PS00097; BTB.1.  
 DR PROSITE; PS00354; HMGI\_Y.1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 6.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
 DR DNA-binding; Metal-binding; Zinc-finger;  
 KW SEQUENCE 641 AA; 69138 MW; C4BEA972C15A877 CRC64;  
 SQ

Query Match	16.4%;	Score 457.5;	DB 11;	Length 641;
Best Local Similarity	24.4%;	Pred. No. 6e-26;		
Matches 165;	Conservative 59;	Mismatches 190;	Indels 263;	Gaps 24

```

0Y 7 REGALGYRREFRHSRSDVJGNINLEBRLR-GILTDVTLVGGQBLRHKXVLJGSGEPYS 65
Db 11 PSGC--YYUQYSRSTEMHNLNQQARKNGRFGCDVLLRVGDESFPHRAVLAAJCEYFES 68
0Y 66 IFRGAVG--VDV--LSLEGP--EARGAPLLDEMYTSRLRL 103
Db 69 VFSAQJGGGAADGGPRADVGAAAAPRGAGGSGRELEMH7ISSKVGDILDFAYTSRIYV 128
0Y 104 SPATPAVLAAATYLOMEHVNOACHRFQASVEPLGISRLPRLAEPTPTA-----P 157
Db 129 RLSEBPELTTAKKFLMNSVIELCOEVIQS-----NVQILVPRADYILFRD 177
0Y 158 PGSRRSSECHPDPPTESRSQSGPPSPASBPDKCMKKYKYLVLNQSQAQSLVGERS 217
Db 178 PGT--SDIGF-----LDMNGAAMANSNGIAG 204
0Y 218 SGQPCPQARLPBGDEDASSSSSSSSSSSEGR-IPRQSRLEPTAATVQFKCAPASTPY 276
Db 205 SMQPEEEAARATGAALAGASLPLVLEGVRLPLVAVG--LSQJLLTSPFPNVASSAP- 260
0Y 277 LLTSQ-----AQDTSGSPS--ERARPLP--GSEFPSCONC--EAVAGCS- 315
Db 261 -LTSRGRGRPRKANULDBMFGSPGGLREBAGILPQGLCGKVFTDARLRLQHEQHGVTL 319
0Y 316 --GLDSLVP-----GBDKYK-----COLCRSSFPYKGNLASHRTVHTG 353
Db 320 QUGYIDLPPRPLGNGELPISEDDPDGKRKSRTKQVACETCGKIFRDVYHLNHNKLSHG 379
0Y 354 EKPYHCSICGARFNRPRANLKTISRJISGE--KPYKCTGCSRVUQRSOPSGFGOKPAR 411
Db 360 EKPYSCPUGLARPKKRDMSYHVRSHDGSVGKRYLQSGCKG----- 422
0Y 412 GGVGQKGGFCSSORODLKSPPSOVAHLRAHV-LIHTGEKPYRCPCTGTRPHLQTLKSHV 470
Db 423 -----SRPDLNHGIQVHTSERPHKQTCNASEPATRDLRLSHL 461
0Y 471 RIH----- 473
Db 462 ACHBDKVPQVCGKYLRALYADHLTKHSBGPSPFCISCNBQOKCSHODLIESSDYG 521
0Y 474 -----TGEKPYHCPCGLFPHRHSQJRLRLRQK 501
Db 522 LSDASDLKTRBKQSANSGFSCDVAVPKNNMESDGKKYPCPEZGSGSPRSSTYLNKHQ-- 579
0Y 502 HGAATNTKHYHIIIGSP 518
Db 580 -----KVHVALRGLP 589

```

## RESULT 10

ID	Q9HBE1	PRELIMINARY;	PRT;	687 AA.
AC	Q9HBE1;			
DT	01-MAR-2001	(TrEMBLrel. 16, Created)		
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
DE	Zinc finger sarcoma gene long C isoform.			
GN	ZSG.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RA	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20404866; PubMed=10949935;			
RA	MacCrangelo T., Modena P., Tornelli S., Bullrich F., Testi A.,			
RA	Mezzelani A., Radice P., Azzarelli A., Pilotti S., Croce C.,			
RA	Pierotti M., Sorzi G.;			

RT "A novel zinc finger gene is fused to EWS in small round cell tumor.";  
 RL OncoGene 19:13799-13804(2000).  
 DR EMBL, AF254085; AACG03034.1; -.  
 DR HSSP; Q05516; IC93.  
 DR InterPro; IPR000637; AT\_hoek.  
 DR InterPro; IPR000515; BFD\_transp.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR000822; ZnF\_C2H2.  
 DR Pfam; PF021178; AT\_hoek; 1.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF00096; Zf-C2H2; 8.  
 DR ProDom; PD000003; ZnF\_C2H2; 1.  
 DR SMART; SMO0384; AT\_hoek; 2.  
 DR SMART; SMO0225; BTB; 1.  
 DR SMART; SMO0355; ZnF\_C2H2; 8.  
 DR PROSITE; PS00402; BFD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE; PS50097; BTB; 1.  
 DR PROSITE; PS00354; HMG\_Y; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 6.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 SQ SEQUENCE 667 AA; 74060 MW; 071C39A78C6D110A CRC64;

Query Match 16.0%; Score 447; DB 4; Length 687;  
Best Local Similarity 26.6%; Pred. No. 3.9e-25;  
Matches 158; Conservative 60; Mismatches 193; Indels 184; Gaps 25;

```

0Y 7 BEGALGVREFTRRSHSDVJLGNINLELR-GLTDTVLVGGQPLBAHKVJLJACSEFFS 65
Db 11 PSGC--YTYQVSRSHSTEMLHNINQOKKONGRCPDVLRLVGDSEFPFHRJVLAAACSYEFES 68
0Y 66 IFRGAGV-----VDV-----LSLPGCP-----EARGFAPLLDEMYTSRLTL 103
Db 69 VFSNAGDGGGAADGGPADVGGAHTAPRGGAGAGSRLEMHNTISSKYPGDIIDPRATYSRIYV 128
0Y 104 SPATAVAVLAATYIQEMHVVQACHRTQASYEPLGISLRPLAEAPTRPTAP-----P 157
Db 129 RLESPEELMTAKFLMRSVIEICQEVKOS-----NVQILVPRAADIPLFRP 177
0Y 158 PGSPRRSGHDPPTRESRC---SGCPSPSPASPDKACMKKKYKTVLNSQASQSGSLV 214
Db 178 PGT--SDUGPLDMDNMGALANSGIAGSNQPEEA-----APPAAMATA 221
0Y 215 ERSSGQCPQOA-RLEPSGDEASIS---SSSSSSSSEEGPIPPQSRSLSPTAATVGEKGA 270
Db 222 GQASLPVLPGVDRLEPMVAGPLSPOLLTPFPFSVASAPRLTKRGGRGRKAN----- 274
0Y 271 PASTPYLLTISOADTSGSPS--ERARPLP-----GSEFFSCONC---EAVAGCS--GLD 318
Db 275 -----LLDSM--FGSPGGLREAGILPCGLGKGVFTDANRLRGHEAOGHTVLSLDGYI 324
0Y 319 SLVP-----GDEDKPYK-----COLCRSSFRTKGNLASHRTYHTGEKFPN 358
Db 355 DLPPRLGLENGLPISEDDPGRPKRSTRKQYACELCGKI FRVUYNLNRKLSHSGCKPYS 384
0Y 359 CSICGARFNRPANLTKHSRISGE--KPYKCECTSGSRFVQVNSOPSPSGQKPARGVQO 416
Db 385 CPVCSLARKRKADRMHYHRSHDGSYKPYICOSCCKGF----- 422
0Y 417 KGGFCSSORODLKSPDSQVAHLRAHV-LIHTEKEVPCPTCGTRFRNHQLTKSHVRIHG 475
Db 423 -----SRPDHLNGHIKQVHTSERDHKKCOTCNASFATRDRLSLHSLACHED 466
0Y 476 EKPYHCDCGLH-----FRKSOQLRLHLROKKG 503
Db 467 KVP--CQVCGKYLRAAYADHLKKGSEGSNFCISICNRPSSASYLKVAHYKTHHG 519

```

## RESULT 11

Q9R161	
ID Q9R161	PRELIMINARY;
AC Q9R161;	PRT;
DT 01-MAY-2000 (TRENBLEL, 13, Created)	703 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
DB 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
DR zinc\_finger protein ZFP235.  
GN ZFP235.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hauser L., York M., Shannon M., Stubbs L.;  
RT "Differential expansion of homologous zinc-finger gene families in  
human chromosome 19q13.2 and mouse chromosome 7.",  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; AF167321; AAD45930.1; -.  
DR HSSP; P25490; 12NM.  
DR MCD; MG11929117; Zfp235.  
DR InterPro; IPR001909; KRAB.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; zf-C2H2; 15.  
DR PRINTS; PR00046; ZINC\_FINGER.  
DR ProDom; PD000003; Znf\_C2H2; 12.  
DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; Znf\_C2H2; 15.  
DR PROSITE; PS50805; KRAB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 15.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 15.  
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.  
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.  
SQ SEQUENCE 703 AA; 79398 MW; 1A1DA00D1DD7386 CRC64;

Query Match 15.9%; Score 442.5; DB 11; Length 703;  
Best Local Similarity 28.5%; Pred. No. 8; e-25;  
Matches 117; Conservative 52; Mismatches 142; Indels 99; Gaps 14;

QY 201 VLNSQASQAGSLVGER---SSGQPCQARLP-----SSDEASSSS 237  
DB 185 ILNCISHQDNHILHKDKVPSSGD-CDQVIFPMACLTQSPVTEOKAYQCNCTQDAFIDS 243  
QY 238 SSSSSSS---SEGRIPGPGSRISPTAATQ-----FKCG-APASPTLLTQ 281  
DB 244 PSETQOVLGNKSPVHGTYEDTSYSSVIGQSVHGRKKRWCEGCKAFSSQSNLQTHQ 303  
QY 282 AQDTSGSP-----SERARLP---GSFFSCNCEAVAGSCSGDLSTV---P 322  
DB 304 RVHTGKPYTCPCPCGSGSFNOSSHLYALHPIHTGKPYCCDNCG--KQFSSTDLNTHCKV 361  
QY 323 GDEDKPYKCOLCRSFRYKGNLASHRTVHTGKPYHCISICGARFNRPAHLKTHSRHSGE 382  
DB 362 HTGKPYKCEVCGKGFQSHLOAHRIHTGKPYKCGDCGKRFSSSNLHTQHVHTE 421  
QY 383 KPKYCETGSRF-----VQVRSPPSGFQKPARGVGQKGGFSSSRDLKSP 431  
DB 422 KPYKCEGCKRFLSFLNLSHQVHTGKPY--YKCEGCKGSSASFSFQSHQVHTGK 478  
QY 432 P-----SQVAHLRAVLHTGKPYKPCPTCGRFRHLQTLSHVRIHTGKPYHC 481  
DB 479 PFCSCVCGKGFSSSFQAHQHVHTGKPYRCDOVCGKRFMSNLNHNQHVHTGKPYK 538  
QY 482 DPCGLFRHKSQQLRLH-----LRQHGATNTKVVYHILG 517  
DB 539 EECGKGFSSQSNLQAHQSVHTGKPYKPCNAQCKRFGQASHLQAHQVHTG 588

RESULT 12

Q9HBE2 PRELIMINARY; PRT; 537 AA.  
AC Q9HBE2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
DE Zinc finger sarcoma gene long B isoform.

GN ZSG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=20404868; PubMed=10949935;  
RA Mascangelo T., Modena P., Tornelli S., Bullrich F., Teati A.,  
RA Mezzelani A., Radice P., Azarelli A., Pilotti S., Croce C.,  
RA Pierotti M., Sozzi G.;  
RT "A novel zinc finger gene is fused to EMS in small round cell tumor.";  
RL Oncogene 19:3799-3804 (2000).  
DR EMBL; AF254084; AAG09033.1; -.  
DR HSSP; Q05516; 1CS3.  
DR InterPro; IPR000637; AT hook.  
DR InterPro; IPR000515; BPD\_transp.  
DR InterPro; IPR000210; BTB\_P02.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF02178; AT hook; 1.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 6.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00384; AT\_hook; 2.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 6.  
DR PROSITE; PS00402; BPD\_TRANS\_P1NN\_MEMBER; UNKNOWN\_1.  
DR PROSITE; PS50097; BTB; 1.  
DR PROSITE; PS00354; HMG1\_Y; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 5.  
KW DNA-binding; Metal-binding; Zinc-finger.  
KW DNA-binding; Metal-binding; Zinc-finger.  
SQ SEQUENCE 537 AA; 57995 MW; 950792413B1F5DB CRC64;

Query Match 15.8%; Score 442; DB 4; Length 537;  
Best Local Similarity 27.3%; Pred. No. 6; e-25;  
Matches 156; Conservative 61; Mismatches 197; Indels 158; Gaps 25;

QY 7 PEGALGYREFFRHSADVGLNLELRRL-GITDVTLLVGGQPLRAHKAVALIACGFFYS 65  
DB 11 PEGC--YTYQVRHSTHMLNINQKRGRCVDLLVGVDSFPAHRAVALACSEYFS 68  
QY 66 IFRGAVG-----VDV---LSLPGP-----EARGFAPLDFTYSRL 103  
DB 69 VSAQIGDGCAGADGADPADVCGATAAPGAGAGSRELEMHITISKVFGLDLPATYSRIIV 128  
QY 104 SPATPAVLAATYIQMEHVQACRFIOASTEPGIGISLRPLAEPPTPTA-----P 157  
DB 129 RLSEPEELMTAKKFLMRSVIEICQEVIKQ-----NVQILVPPARADIMLFRP 177  
QY 156 PGPSPRESEHPPEPSRSC---SQGPPSPAPDPKACNMKKYKIVLNSQASQAGSLVG 214  
DB 178 PGT--SDGFLPDMTNGALANSNGINGSMQPEEA-----ARAAGAAIA 221  
QY 215 ERSQGPCQA-RLPSGDEASS---SSSSSSSEEGPIPGQRLSPATAVQFKCGA 270  
DB 222 GQASLPVLPVGVRLPMVAGPLSPQLTSPFVSASAPPLTKRGGRBRKAN----- 274  
QY 271 PASTYLLTQAQDTSGSPS--ERARPLP---GSEFFSCQNC---EAVAGCS--GLD 318  
DB 275 -----LDSM---FGSPGGLREAGILPCGLCGVFYFDANRLRQHEKOHGVTSIQIGI 324  
QY 319 SLVP-----GDEDKPYK-----COLCRSFRYKGNLASHRTVHTGKPYH 358  
DB 325 DLPPRLGNGCLPISEDPDGPRKSRTRKQVACEICGKIFRDVYLNHKKLSHSEKPY 384  
QY 359 CSICGARFNRPAHLKTHSRHSGE--KPYKCYTCGRFVQVRSOPSGFQKPARGVGQ 416  
DB 385 CPVCGLFRKRDMSYVHRSHDGVSKPYICQSGCKG----- 422  
QY 417 KGGFGSSORODLKSPSQVAHLRAHV-LHTGKPYKPCPTCGTRRHLQTLKSHVRIHTG 475  
DB 423 -----SRPDHLNGHIKQVHTSERPHKQTCQCNASPATRDLRSLHLCHE 466



QY 476 EKRYHCDPCGLHFRHKSQRLHROKIGATN 507  
 DB 467 KVP--CQVCGKTLR-AAVADHLKKHSESPN 495

## RESULT 13

Q9PVP8 PRELIMINARY; PRT; 470 AA.  
 ID 09PVP8  
 AC 09PVP8  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Champignon.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21255550; PubMed=11357190;  
 RA Goto T., Hasegawa K., Kinoshita T., Kubota H.Y.;  
 RT "A novel POZ/zinc finger protein, Champignon, interferes with  
 RT gastrulation movements in Xenopus";  
 RL EMBL, AB029074; BAA87059.1; -.  
 DR HSSP; P08047; 1SP2.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR000822; Znf\_C2H2.  
 DR Pfam; PFO0651; BTB; 1.  
 DR Pfam; PFO0096; zf-C2H2; 4.  
 DR SMART; SM00355; Znf\_C2H2; 4.  
 DR SMART; SM00097; BTB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_3.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
 DR DNA-binding; Zinc-finger.  
 KW DNA-binding; Zinc-finger.  
 SQ SEQUENCE 470 AA; 559121688295FD CRC64;

Query Match 15.8%; Score 441; DB 13; Length 470;  
 Best Local Similarity 26.6%; Pred. No. 7e-25;  
 Matches 138; Conservative 57; Mismatches 184; Indels 140; Gaps 13;

QY 17 FTRHSDVGNINELRLGILTDVTLVGGQPLRAKAVLIACSGFFYSIFRBAVG-- 74  
 DB 13 FTRHSDVGNINELRLGILTDVTLVGGQPLRAKAVLIACSGFFYSIFRBAVG-- 72  
 QY 75 ----VDVLSLPGPEARGAPLLDFMYTSRLSPATAVAVLAATYLMHEHYOCHR 129  
 DB 73 PDVCGDPLK---PALG--ALDDPATYATLTISGNMMDVRAARLLEPCVAVACVD 126  
 QY 130 FIOAS--YEPGLISLRPLEAEPPTPTAPPSPRSEGHDPPTSRSCSQCPSPASP 187  
 DB 127 ILOCHREMGDADLECF-----LBARQVLSYMGENVATPPPAESP 173  
 QY 188 DPACMWMKKYKTVLNSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSSSSSSSE 247  
 DB 174 PPHPHN-----IPVPPKSVQIIIPRRGRKFLQVNPFRRIQNNLIRGADSLERDASHA 227  
 QY 248 GPILPGQSRSLPTAATVQFKG---APASTPYLLTSQADTSG---SPSERARPLPSG 299  
 DB 228 GSPPNPPLSGEYEVADNGLGHTIFVPPSPREIISD--HETSDMGPHNPYLENV--- 283  
 QY 300 EPPSCNCEAVAGCGSLDVLPGDBDK-----PYICQLCRSSFRYKGLASHRTV 350  
 DB 284 ----SAGLDV-----SDKLVKRRSQLPQRCPCVCHKIIGHAGRLPRHMT 324  
 QY 351 HTEKPYHSGICGARFRPRLNLTSHRISGEKPYCETGSGRFVQVGRSPSPGFGKXA 410  
 DB 325 HTEKPPVCCVCCVTRNDKLIHMKHGERPYCEHCSARFL----- 369  
 QY 411 RGVGKGFGFCSSORODLSPSOVALRAHVLHTEKPYPCPTGTRFRLQLTLKSHV 470

DB 370 -----HSDYLNKM 378  
 QY 471 RHTEKPYHCDPCGLHFRHKSQRLHROKIGATNTK 509  
 DB 379 HLHTGRPECSLCHKAFKEDHLORHMKQNCLEVRTR 417

## RESULT 14

Q9QZ48 PRELIMINARY; PRT; 569 AA.  
 ID 09QZ48  
 AC 09QZ48  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Osteoclast-derived zinc finger (OCZF) protein.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE=99408739; PubMed=10477728;  
 RA Kukita A., Kukita T., Ouchida M., Maeda H., Yatsuki H., Kohashi O.;  
 RT "Osteoclast-derived zinc finger (OCZF) protein with POZ domain, a  
 RT possible transcriptional repressor, is involved in  
 RT osteoclastogenesis";  
 RL Blood. 94:1987-1997 (1999).  
 DR EMBL; D88450; BAA6393.1; -.  
 DR HSSP; P08046; 1A1H.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR000822; Znf\_C2H2.  
 DR Pfam; PFO0651; BTB; 1.  
 DR Pfam; PFO0096; zf-C2H2; 4.  
 DR SMART; SM00225; BTB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 3.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_3.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
 DR DNA-binding; Zinc-finger.  
 KW DNA-binding; Zinc-finger.  
 SQ SEQUENCE 569 AA; 60543 MW; C111AB303B0D0FD CRC64;

Query Match 15.8%; Score 440.5; DB 11; Length 569;  
 Best Local Similarity 25.2%; Pred. No. 9.5e-25;  
 Matches 146; Conservative 57; Mismatches 186; Indels 191; Gaps 17;

QY 8 EGALGYVEFTTRHSDVGNINELRLGILTDVTLVGGQPLRAKAVLIACSGFFYSIF 67  
 DB 6 DGPGLI--PFPDHSDDLGLNEGRTOGLCDVILVGRREPTSRVLAACSGYFKLF 63  
 QY 68 RRAVG-----VDVLSLPGPEARGAPLLDFMYTSRLSPATAVAVLAATYLMHE 121  
 DB 64 TSGAVALDOONVEIDFVS-----AEALTALMDFYVTLVSTANVGDIISAALLEIP 117  
 QY 122 HVVQACHRFIOASYPLGISLRPLEAEPPTPTAPPSPRSEGHDPPTSRSCSQCP 181  
 DB 118 AVSRVCTDLLE-----RQIIAADVDGAGPDGA-----CPTDQRULRLRK 158  
 QY 182 -----PSPASDPKACMWMKKY-----KTVLNSQASQAGSLVGRSSGQ 221  
 DB 159 EYLFEPFRSNPNMSLP-PTAFQWPGFSAPDDLDATKEAVAAVAALVAGDCGLPFGP 217  
 QY 222 CPQARLPSGDEASSSSSSSSSSSSSEGLPG---PQRLSTATAVQFKG--APASTPY 276  
 DB 218 PPADRPPTGDEGEGSTGLWPERDEDAIPGGLFPPT--APPATYQNGHYRAGAST-- 273  
 QY 277 LITQADOTSGSPSRARPLPG--SEFFSCNCEAVAGCGSLDVL----- 321  
 DB 274 -----GEEBAVALSE--AAPPEGDSGFLSGAAGEDGDAVDGLAATLLOMMSSVGR 327  
 QY 322 PGDEDKP-----YKQLCRSSFRY 340

Db 328 AGSDSESRPDDKGVMDVYLKTFSGAEGDVPWMSQGEKKIRAKAFQKPICEKVIQ 387  
QY 341 KGNLASHRTVHTGEKPYHCSI CGARFNRPAHLKTHSRHSGEKPYKCEGSHFVQVRQ 400  
Db 388 AGULPHRIHTHTEKPYECNICRRTFRQDKLVHNRKHTGEKPYLQOQGAFA----- 442  
QY 401 PPSGFOGKPARGVGQKGFSSORODLKSPPSOVALRAHVLHTHGEKPYPCPTGTRP 460  
Db 443 ----- 442  
QY 461 RHLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQRLHLRQ 500  
Db 443 -HNYDLKNHMRVHTGLRPYQDCSCKTFVNSDHLHRLKK 481  
RESULT 15  
ID 09H17 PRELIMINARY; PRT; 545 AA.  
AC 09H17;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE D0322G13.2.1 (zinc finger protein FLJ21794, isoform 1).  
GN D0322G13.2.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lovell J.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL096677; CAC17422.1; -  
DR HSSP: P07248; 1ARF.  
DR InterPro: IPR000210; BTB\_POZ.  
DR InterPro: IPR000822; Znf\_C2H2.  
DR Pfam: PF00651; BTB; 1.  
DR Pfam: PF00096; zf-C2H2; 8.  
DR ProDom: PD000003; Znf\_C2H2; 1.  
DR SMART: SM00325; BTB; 1.  
DR SMART: SM00355; Znf\_C2H2; 8.  
DR PROSITE: PS50097; BTB; 1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 8.  
KW DNA-binding; Metal-binding; Zinc-finger.  
SQ SEQUENCE 545 AA; 61982 MW; AAA9820E319C361E CRC64;  
Query Match 15.8%; Score 439.5; DB 4; Length 545;  
Best Local Similarity 25.2%; Pred. No. 1.1e-24;  
Matches 149; Conservative 61; Mismatches 200; Indels 181; Gaps 19;  
QY 23 DVLGNINELRLRGLDVTLLVGQPLR---AHKAVLIACSGPFYSIFRGRAGV--GVDV 77  
Db 16 NLHHEHHEHLRLHGLCDVTLSVEYQGRKDFMAHKAVALATSKFEKVFLEKSVDTGTRT 75  
QY 78 LSLPGGEAAGAPALDLFWYTSRLSPATAPAVLAATYLOMEHVVOACHRFIOASYEP 137  
Db 76 NVYLNEVQVADFPASPLEFYTAQVIEDRVQRMLEVAEKLKCLDLSSETCFOLKQMLSS 135  
QY 138 LGISLRLP-----EAEPPT---PPTAPPSSFRSEGHHPDPTES-----RSCSGQPPSP 184  
Db 136 VLLLELONFSESQEVESSSGSQVSAAPRASVATDGPHPGLTDSLDPGERASNGMSSD 195  
QY 185 ASDPDKACNMKKYKYIV-----LNSQASQAGSLV 213  
Db 196 LPPKSKDKLDKKKEVVKPPYKIRASGRLAGRKVFVEIPKKKYTRRLREQQKTAEGDV 255  
QY 214 GERSSGQPCPOARLPSS-----DEASSSSSSSSSSSSSECGPIPGP 253  
Db 256 GD-----YRCFQDQSPDRVGTGEMEQVSKNECGQAGALEBELSKKAGPBEHEEBDEBEG 311  
QY 254 QSRLSPTAATVQFKC-----GAPASTPYLLTSAQDTSSGP--- 289

Db 312 KKKSN-----FKSICEKAFLEYKSPFKSHKRGVATEVVYRC-----DTCGQTPAN 359  
QY 290 -----SEBARPLPGSE-FPSCQNC-----EAVAGCSGSLDVLPGDEDEKPYKQCL 333  
Db 360 RGNLKHQRAVHSSSRHFPCELCGKKFKKQDVKNHVLQVHEG-----GSE--RHRGQ 411  
QY 334 CRSSFRYKGNLASHRTVHTGEKPYHCSI CGARFNRPAHLKTHSRHSGEKPYKCEGSHF 393  
Db 412 CGKGLSSKTLALHRTHTGDRPYGCTEGGARFSQPSALKTMRHTHGEKPYFVCDCEGAR 471  
QY 394 FVQVNSQPPSGFGQKPARGVGQKGFSSORODLKSPPSOVALRAHVL-----HTGE 448  
Db 472 FTQ-----NHMLYHNRCHTGE 488  
QY 449 KPYPCTGTRPRLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQRLHLR 499  
Db 489 RFPKCEGCKSPASKKEYLKHNRHTTGSKPFCEVCPTFAORNSLYQHIX 539

Search completed: July 14, 2003, 18:20:56  
Job time : 27.0314 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 ; Search time 71.3175 Seconds  
(without alignments)  
3843.329 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730  
Sequence: 1 FCLQGRVWLRENGQHPST.....MIVKKRYSTTRSSAQSSR 2057

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.Geneseq.101002.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10730	100.0	2057	22	AAE11890
2	10626	99.0	2058	22	ABB97219
3	10247	95.5	2048	22	AAE11891
4	7166.5	66.8	2408	22	ABG10631
5	6775.5	63.1	1515	22	ABG10624
6	5385	50.2	1031	22	AAU17082
7	5166	48.1	984	22	AAW93762
8	5146.5	48.0	1442	22	ABG25601
9	4450	41.5	869	21	AAW42020
10	3238.5	30.2	713	22	ABG10632

11	2032	18.9	394	22	AAE93020	Human protein sequ
12	1947	18.1	402	22	AAU23129	Novel human enzyme
13	1843.5	17.2	2167	22	ABB60369	Drosophila melanog
14	1789	16.7	348	22	AAU17508	Novel signal trans
15	1705	15.9	2129	22	ABB62828	Drosophila melanog
16	1495	13.9	2424	22	ABB58924	Drosophila melanog
17	1492	13.9	2548	20	AAV05781	Human myosin IXa
18	1476	13.8	1544	21	AAE48638	Arabidopsis thalia
19	1466.5	13.7	1493	21	AAE48639	Arabidopsis thalia
20	1450.5	13.5	1556	21	AAE30438	Arabidopsis thalia
21	1444.5	13.5	1483	21	AAE48640	Arabidopsis thalia
22	1441	13.4	1505	21	AAE30439	Arabidopsis thalia
23	1434.5	13.4	697	17	AAW00603	Human non-conventi
24	1426	13.3	631	17	AAW00604	Human myosin VII
25	1419	13.2	1495	21	AAE30440	Arabidopsis thalia
26	1372.5	12.8	1792	22	ABB71108	Drosophila melanog
27	1352	12.6	1839	21	AAE94291	Caenorhabditis ele
28	1344.5	12.5	301	22	ABG25592	Novel human diagno
29	1322.5	12.3	1972	17	AAW00024	Smooth muscle myos
30	1313	12.2	2056	22	ABB59344	Drosophila melanog
31	1308	12.2	1453	22	AAE39213	Human polypeptide
32	1306	12.2	1469	22	AAE39214	Human polypeptide
33	1305.5	12.2	1754	21	AAE52410	Arabidopsis thalia
34	1305.5	12.2	1769	21	AAE52409	Arabidopsis thalia
35	1305.5	12.2	1804	21	AAE52408	Arabidopsis thalia
36	1303	12.1	1988	22	AAW40999	Human polypeptide
37	1303	12.1	1988	22	AAW40999	Human polypeptide
38	1297.5	12.1	1960	22	AAW78854	Human protein SER
39	1297.5	12.1	1963	22	AAW78838	Human alpha-myosin
40	1274	11.9	1939	23	ABB77096	Human alpha-myosin
41	1269	11.8	1948	22	ABG21233	Novel human diagno
42	1268	11.8	1879	22	AAE25750	Human protein sequ
43	1262.5	11.8	1120	21	AAE94292	Helianthus annuus
44	1261	11.8	2067	22	ABB71125	Drosophila melanog
45	1249.5	11.6	1857	23	AAU84350	Protein MYH1 diff

#### ALIGNMENTS

RESULT 1						
AAE11890	AAE11890	standard; Protein; 2057 AA.				
XX	XX					
AC	AAE11890;					
DT	18-DEC-2001 (first entry)					
DE	Angiogenesis associated human myosin XI (hMX1) protein.					
XX	XX					
KW	Angiogenesis associated protein; AAP; cytosaratic; cardiant; gene therapy;					
KW	ophthalmological; valvultery; myocardial infarction; macular degeneration;					
KW	diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;					
KW	rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;					
KW	cancer; therapeutic; diagnostic; human; myosin XI; MX1.					
OS	Homo sapiens.					
XX	XX					
PN	WO200170808-A2.					
XX	XX					
PD	27-SEP-2001.					
XX	XX					
PF	22-MAR-2001; 2001WO-US09609.					
XX	XX					
PA	22-MAR-2000; 2000US-191134P.					
XX	XX					
PA	(CURA-) CURAGEN CORP.					
XX	(GUTH) GENENTECH INC.					
XX	XX					
PI	Rastelli LK, Gerritsen M;					
XX	XX					
DR	WPI; 2001-602775/68.					
XX	XX					
DR	N-PSDB; AAD19119.					

XX Novel angiogenesis associated polypeptides and polynucleotides encoding  
PT the polypeptides, useful for modulating angiogenesis and for treating  
PT tumors and cancers -  
XX  
PS Claim 1, Page 31-38, 159pp; English.  
XX  
CC The invention relates to angiogenesis associated proteins (AAP) and their  
CC corresponding cDNA molecules, which are useful for modulating  
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting  
CC wound healing, for example after organ transplantation, and in the  
CC treatment of tumours, myocardial infarction, cancers, diabetic  
CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
CC AAP proteins and DNA's are useful in potential prophylactic and  
CC therapeutic applications implicated in a variety of disorders including  
CC those related to angiogenesis, and also in diagnostic applications.  
CC AAP cDNA is also useful in gene therapy. The invention also relates to  
CC a method for screening a tissue sample for tumorigenic potential. AAP  
CC proteins are used to screen drugs or compounds that modulate AAP activity  
CC or expression as well as treating disorders characterised by insufficient  
CC or excessive production of AAP or production of AAP forms that have  
CC decreased or aberrant activity compared to the wild type protein, or  
CC modulate biological function that involve AAP. The present sequence  
CC is human myosin XI (hmxi) protein which is an angiogenesis associated  
CC protein (AAP) of the invention. Hmxi is upregulated in the in vitro model  
CC of angiogenesis and is likely to be involved in transducing signals from  
CC angiogenic factors, perhaps modulating the cytoskeleton.  
XX  
SQ Sequence 2057 AA;  
Query Match 100.0%; Score 10730; DB 22; Length 2057;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2057; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FCHGTRVWLRKNGQHPSTVNSCAGIYVETDYGQVFTYKOSTTHOKVTMHPNNE 60  
DB 1 FCHGTRVWLRKNGQHPSTVNSCAGIYVETDYGQVFTYKOSTTHOKVTMHPNNE 60  
QY 61 GVDNMSLTLEHSGIMYNLFRYKRNQWITYGSLASVNPYQPLAGLEPAPMEQYSR 120  
DB 61 GVDNMSLTLEHSGIMYNLFRYKRNQWITYGSLASVNPYQPLAGLEPAPMEQYSR 120  
QY 121 RHGELRPPIFAIANECYRCMLKRNHNOCLLIGESGAGTSTKLLKLSYISQOSLE 180  
DB 121 RHGELRPPIFAIANECYRCMLKRNHNOCLLIGESGAGTSTKLLKLSYISQOSLE 180  
QY 121 RHGELRPPIFAIANECYRCMLKRNHNOCLLIGESGAGTSTKLLKLSYISQOSLE 180  
DB 121 RHGELRPPIFAIANECYRCMLKRNHNOCLLIGESGAGTSTKLLKLSYISQOSLE 180  
QY 181 LSLKEKTSYCEBAILLESSPIWEAFGNAKTYNNNSRFGKFOVLNLCQKNTIGSGRIYDC 240  
DB 181 LSLKEKTSYCEBAILLESSPIWEAFGNAKTYNNNSRFGKFOVLNLCQKNTIGSGRIYDC 240  
QY 241 ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREPEFLSTPENYHYLNSGCEVEDTISD 300  
DB 241 ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREPEFLSTPENYHYLNSGCEVEDTISD 300  
QY 301 QESFREYITMDWQSKSEVRSRLLAGILHGNIEFTTAGAQSFTALGRSABELL 360  
DB 301 QESFREYITMDWQSKSEVRSRLLAGILHGNIEFTTAGAQSFTALGRSABELL 360  
QY 301 QESFREYITMDWQSKSEVRSRLLAGILHGNIEFTTAGAQSFTALGRSABELL 360  
DB 301 QESFREYITMDWQSKSEVRSRLLAGILHGNIEFTTAGAQSFTALGRSABELL 360  
QY 361 GLDPTQTDALTORSMFLRGEELITPLNQAVDSRSLAMALYACCFEVIKIKISRIK 420  
DB 361 GLDPTQTDALTORSMFLRGEELITPLNQAVDSRSLAMALYACCFEVIKIKISRIK 420  
QY 421 GNEDFSGISGLIDFGEFENFEVNHFEQFINYANEKLOEYFNKHI FSLSEQLRESEGLVWE 480  
DB 421 GNEDFSGISGLIDFGEFENFEVNHFEQFINYANEKLOEYFNKHI FSLSEQLRESEGLVWE 480  
QY 481 DIRMINDGECLDIIEKKLGLAIALINESHPQATDSTLLEKLSQHANNHFYKPRVAVN 540  
DB 481 DIRMINDGECLDIIEKKLGLAIALINESHPQATDSTLLEKLSQHANNHFYKPRVAVN 540  
QY 541 NFGVKYHAGEVOYDVNGILEKNDTPFDLNLNLRSPDPFIYDLFEHVSRRNQDTLKC 600  
DB 541 NFGVKYHAGEVOYDVNGILEKNDTPFDLNLNLRSPDPFIYDLFEHVSRRNQDTLKC 600

QY 601 GSKHRRPTVSQPKVDSLHSLMATLSSNPFVRCIKPMNQMPQFDQAVVLNQLRYSG 660  
DB 601 GSKHRRPTVSQPKVDSLHSLMATLSSNPFVRCIKPMNQMPQFDQAVVLNQLRYSG 660  
QY 661 MLETVAIRKAGVAVRRPFDQFYKRYKVLNRNLALBEDVRGKCTSLQLDYDASNEMQLK 720  
DB 661 MLETVAIRKAGVAVRRPFDQFYKRYKVLNRNLALBEDVRGKCTSLQLDYDASNEMQLK 720  
QY 721 TYVFLRESLEQLEKRESEVSHAMVIRAHVIGFLARQKRYVCVYIQKNYRAFL 780  
DB 721 TYVFLRESLEQLEKRESEVSHAMVIRAHVIGFLARQKRYVCVYIQKNYRAFL 780  
QY 781 RRRFLHKKAAIVFOKOLRGQJARRVYRLAEKREBEKKKQSEBEKKKREERERER 840  
DB 781 RRRFLHKKAAIVFOKOLRGQJARRVYRLAEKREBEKKKQSEBEKKKREERERER 840  
QY 841 ERREAEALRQSEBEKTRKOELALQSKQAEALTELEKQKNKQVEILRLKEIEDLQ 900  
DB 841 ERREAEALRQSEBEKTRKOELALQSKQAEALTELEKQKNKQVEILRLKEIEDLQ 900  
QY 901 RKMEQOELSLTASLOKLERDOELRLREBACRAOEFLESLNPFDEICVRNIERSL 960  
DB 901 RKMEQOELSLTASLOKLERDOELRLREBACRAOEFLESLNPFDEICVRNIERSL 960  
QY 961 SGSEFSSSLASACEKENFNSQPYPEEYDEGFEADDAFKDSPNESEHSDQRTS 1020  
DB 961 SGSEFSSSLASACEKENFNSQPYPEEYDEGFEADDAFKDSPNESEHSDQRTS 1020  
QY 1021 GRTSPDSEEDPYNDIVPTSPSADSTVLLAPVQDSGLHNSSGSSTYCMQONMGD 1080  
DB 1021 GRTSPDSEEDPYNDIVPTSPSADSTVLLAPVQDSGLHNSSGSSTYCMQONMGD 1080  
QY 1081 LPSPGDYYDDDDYEDGALITSGSSVTFSNSGQMSPDYRCSVGTYNSSGAYRESSGA 1140  
DB 1081 LPSPGDYYDDDDYEDGALITSGSSVTFSNSGQMSPDYRCSVGTYNSSGAYRESSGA 1140  
QY 1141 QGSFEDSEEDPDRPTDDELSTYRDSVYSCTLLEPHSFLYMKGLMSWRRKRCVLKD 1200  
DB 1141 QGSFEDSEEDPDRPTDDELSTYRDSVYSCTLLEPHSFLYMKGLMSWRRKRCVLKD 1200  
QY 1201 EFTFLPFRSQEALKQGLHKKGGSTLSRBWKKRWPLRQSKLWYFENDSEELKGV 1260  
DB 1201 EFTFLPFRSQEALKQGLHKKGGSTLSRBWKKRWPLRQSKLWYFENDSEELKGV 1260  
QY 1261 EYRTAKEIIDNTTKENGIDIIADRTPHLIAESPDAQSWFVLSQVHASTDQEIQEMHD 1320  
DB 1261 EYRTAKEIIDNTTKENGIDIIADRTPHLIAESPDAQSWFVLSQVHASTDQEIQEMHD 1320  
QY 1321 EQANPQNAVGLDVGILDSVCSDSPDRPNSFVITIANRVLHCNADTEPEMHWITLLQR 1380  
DB 1321 EQANPQNAVGLDVGILDSVCSDSPDRPNSFVITIANRVLHCNADTEPEMHWITLLQR 1380  
QY 1381 SKGDRRVGEQEFIVGMLHKEVNSPKMSLSLKKRWPLTNSLDYKSSSKMLKIGT 1440  
DB 1381 SKGDRRVGEQEFIVGMLHKEVNSPKMSLSLKKRWPLTNSLDYKSSSKMLKIGT 1440  
QY 1441 LVINSLCSVVPDEKIFKETGYMNTYVGRKHCYLLYTKLNLBAIRMSVIONVDTKAP 1500  
DB 1441 LVINSLCSVVPDEKIFKETGYMNTYVGRKHCYLLYTKLNLBAIRMSVIONVDTKAP 1500  
QY 1501 IDPTQQLIODIKENCLNSDVVEQIYKGNPILRYTHHPHLSPLPLPYGDINLNLKDG 1560  
DB 1501 IDPTQQLIODIKENCLNSDVVEQIYKGNPILRYTHHPHLSPLPLPYGDINLNLKDG 1560  
QY 1561 YTTLODEAKITNSIQOLESMSDPIPIIQGILQTHDLRPLDELYCQLIKQTNVPHHG 1620  
DB 1561 YTTLODEAKITNSIQOLESMSDPIPIIQGILQTHDLRPLDELYCQLIKQTNVPHHG 1620  
QY 1621 SVGNLYSMQILTCLCTFLPSRGIKYLKFKHLKRIREQPGTEMEKVALFTYESLJKTGC 1680  
DB 1621 SVGNLYSMQILTCLCTFLPSRGIKYLKFKHLKRIREQPGTEMEKVALFTYESLJKTGC 1680

QY 1681 REFVPSRDEIEALHROEMTSTVYCHGGSCKITINSHTAGVEVKLRGLAMEDSRNM 1740  
DB 1681 REFVPSRDEIEALHROEMTSTVYCHGGSCKITINSHTAGVEVKLRGLAMEDSRNM 1740  
QY 1741 PALFENGHVDAKISRRTVAVALAKPEKLAATSEVGDLEFKPFYFLVYCFLLTDVNPXDS 1800  
DB 1741 PALFENGHVDAKISRRTVAVALAKPEKLAATSEVGDLEFKPFYFLVYCFLLTDVNPXDS 1800  
QY 1801 VEFAPMEQAHBAVHGHHPABENLQVLAALRLOGLCYTLTHAALPPLSEVYSLORLK 1860  
DB 1801 VEFAPMEQAHBAVHGHHPABENLQVLAALRLOGLCYTLTHAALPPLSEVYSLORLK 1860  
QY 1861 ARISSQSTKFTPCERLEKERTSFLEGTLLRRSFRGVSVAQKVEEOMLDMWKEEVSAR 1920  
DB 1861 ARISSQSTKFTPCERLEKERTSFLEGTLLRRSFRGVSVAQKVEEOMLDMWKEEVSAR 1920  
QY 1921 ARIIDKRRKFGQMNQOQAMAKYMALIKENPGYSTLPDVECKEGGFROPLMGVSADVS 1980  
DB 1921 ARIIDKRRKFGQMNQOQAMAKYMALIKENPGYSTLPDVECKEGGFROPLMGVSADVS 1980  
QY 1981 VKRGGGRPLVEFOYEHILSPGAPLANTYKIVDERELLETSEVVDVAKLMKAYISMTV 2040  
DB 1981 VKRGGGRPLVEFOYEHILSPGAPLANTYKIVDERELLETSEVVDVAKLMKAYISMTV 2040  
QY 2041 KRRYSTTRGASSQSSSR 2057  
DB 2041 KRRYSTTRGASSQSSSR 2057

RESULT 2  
ABB97219  
ID ABB97219 standard; Protein; 2058 AA.

AC ABB97219;

DT 27-JUN-2002 (first entry)

DB Novel human protein SEQ ID NO: 487.

Human; antiflaemetic; vulnerary; antiinflammatory; immunomodulator;  
antiinfectility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
neuroprotective; antiParkinsonian; protein therapy; EST;  
expressed sequence tag.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

PA (HYSB-) HYSBQ INC.

PI Tang YF, Liu C, Zhou P, Adundi V, Zhang J, Zhao QA, Ren F;  
Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI; 2002-292408/33.

DR N-PSDB; ABB32405.

PT An isolated polynucleotide for treating diseases associated with its  
encoded polypeptide such as cancer and multiple sclerosis -

XX Example 2; SEQ ID NO 487; 509pp; English.

XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 2058 AA;

Query Match 99.0%; Score 10626; DB 23; Length 2058;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2043; Conservative 4; Mismatches 8; Indels 2; Gaps 2;

QY 1 FCLQSTRVMBNGHHPSTVNSCAEGIVRTDYGQYFTYKOSTITHTQKTAMPTNE 60  
DB 4 FTEGTRVMBNGHHPSTVNSCAEGIVRTDYGQYFTYKOSTITHTQKTAMPTNE 63  
QY 61 GVDMAASLTSLHGGSIMNLFORYRNQIMTYIGSILASVNPYOPDLAGYBPATHEOYSR 120  
DB 64 GVDMAASLTSLHGGSIMNLFORYRNQIMTYIGSILASVNPYOPDLAGYBPATHEOYSR 123  
QY 121 RHLGELPHIPALANECYRCMLKRDHNCILIKSGGAGKTESTLILKPLSVISQGSLE 180  
DB 124 RHLGELPHIPALANECYRCMLKRDHNCILIKSGGAGKTESTLILKPLSVISQGSLE 183  
QY 181 LSLKETSCEBRALIESSPIMEAFGNACTVYNNSSRFGKTVOLNICOXNIOGGRIVDC 240  
DB 184 LSLKETSCEBRALIESSPIMEAFGNACTVYNNSSRFGKTVOLNICOXNIOGGRIVDC 243  
QY 241 ILSQNRVVRONPGERNHIFVALAGLEHEREREFTYSTEENHYLNQSGVEDKTSID 300  
DB 244 LL-EKORVRONPGERNHIFVALAGLEHEREREFTYSTEENHYLNQSGVEDKTSID 302  
QY 301 QESFEVITAMDVQFSKEEVEVSRILAGILHLGNIETAGGAQVSPKALAGSABL 360  
DB 303 QESFEVITAMDVQFSKEEVEVSRILAGILHLGNIETAGGAQVSPKALAGSABL 362  
QY 361 GLDPQULTDALTONSMPLRGEILTPLVNQOAVDSRDSLAMALYACCEBMYIKKINSIK 420  
DB 363 GLDPQULTDALTONSMPLRGEILTPLVNQOAVDSRDSLAMALYACCEBMYIKKINSIK 422  
QY 421 GNEDKTSIGIILDFEENFEVNHFEQFINVANELOEYFMKHIFSLERLERSRGLWE 480  
DB 423 GNEDKTSIGIILDFEENFEVNHFEQFINVANELOEYFMKHIFSLERLERSRGLWE 482  
QY 481 DIDWIDNGECUDLEKKGILALINEESHFPQATDSTLEKLSQHANNHFFVYKRVAVN 540  
DB 483 DIDWIDNGECUDLEKKGILALINEESHFPQATDSTLEKLSQHANNHFFVYKRVAVN 542  
QY 541 NFGVHYAGEVQYDVRGILEKNDTFRDILNLLRESFDFIYDLFEHVSSRNNQDTLKC 600  
DB 543 NFGVHYAGEVQYDVRGILEKNDTFRDILNLLRESFDFIYDLFEHVSSRNNQDTLKC 602  
QY 601 GSKHRPTVSSQFKDLSHSLMATLSSNPFFVCIKNNKMPQFOQAVVNLNLRISG 660  
DB 603 GSKHRPTVSSQFKDLSHSLMATLSSNPFFVCIKNNKMPQFOQAVVNLNLRISG 661  
QY 661 MLETYRIKAGYAVRPFQDFYKRYKVMRLNALPEDVRGKCTSLCLYDASNSSEMOLGK 720  
DB 662 MLETYRIKAGYAVRPFQDFYKRYKVMRLNALPEDVRGKCTSLCLYDASNSSEMOLGK 721  
QY 721 TKVFLRBSLEQLEKGRREEVSHAAVIRAHVGLFARKOYRKVLVCYVIIIOKNRAFL 780  
DB 722 TKVFLRBSLEQLEKGRREEVSHAAVIRAHVGLFARKOYRKVLVCYVIIIOKNRAFL 781  
QY 781 RRRFLHLKKAIVPQKQIARVYVQQLAEXRQEQKQKQKQKQKQKQKQKQKQKQKQK 840  
DB 782 RRRFLHLKKAIVPQKQIARVYVQQLAEXRQEQKQKQKQKQKQKQKQKQKQKQKQK 841  
QY 841 ERREALEAQQEESTRKOQELALQSQKEALETRLEKQKQKQKQKQKQKQKQKQKQKQK 900  
DB 842 ERREALEAQQEESTRKOQELALQSQKEALETRLEKQKQKQKQKQKQKQKQKQKQKQK 901  
QY 901 RKKEQOELSLTBASIQKQERDDELRLLEBACRAAOEPLSLNFDIIBCVNIRSL 960

Db 902 RMKEQDELSTLEASLQKLOERRDQELRLLEEACRAAQEFLSESLNDEIDEVNRTERSL 961  
 QY 961 SGGSFSESLAEASACEEKNFNFNSQPYPEEVEVGEFADDAFKDSPNPEHSGDQRTS 1020  
 Db 962 SVGSFSESLAEASACEEKNFNFNSQPYPEEVEVGEFADDAFKDSPNPEHSGDQRTS 1021  
 QY 1021 GRTSDSDSSEEDPYMNDTVPTSPSADSTVLLAPSVODSGSLHNSSSGSESTYCPONAGD 1080  
 Db 1022 GRTSDSDSSEEDPYMNDTVPTSPSADSTVLLAPSVODSGSLHNSSSGSESTYCPONAGD 1081  
 QY 1081 LPSPDGDYDODDYEDGALTSSGSVTFNSYSGSOMSPDRGCVGTNSGAYRFSSEGA 1140  
 Db 1082 LPSPDGDYDODDYEDGALTSSGSVTFNSYSGSOMSPDRGCVGTNSGAYRFSSEGA 1141  
 QY 1141 QSSFEDESEEDFDSRFTDDELSTYRDSVYSCVTLPYFHSFLYKKGGLMWMKRWCVLKD 1200  
 Db 1142 QSSFEDESEEDFDSRFTDDELSTYRDSVYSCVTLPYFHSFLYKKGGLMWMKRWCVLKD 1201  
 QY 1201 ETLWFRSKQBALKQGLHKKGGSSSTLSRRNKKRWFLRQSKLWYFENDSEKLGTV 1260  
 Db 1202 ETLWFRSKQBALKQGLHKKGGSSSTLSRRNKKRWFLRQSKLWYFENDSEKLGTV 1261  
 QY 1261 EVRTAKEIIDNTKENGIDIIIMADRTPHLLAESPEASQWFSVLSQVHASTDDEIQEMHD 1320  
 Db 1262 EVRTAKEIIDNTKENGIDIIIMADRTPHLLAESPEASQWFSVLSQVHASTDDEIQEMHD 1321  
 QY 1321 EQANPQNAVGLDVLGIDSVCASDSPDRPNSFVYITANRVLHGNADTPEEMHMITLLQR 1380  
 Db 1322 EQANPQNAVGLDVLGIDSVCASDSPDRPNSFVYITANRVLHGNADTPEEMHMITLLQR 1381  
 QY 1381 SKGDTVEGGEFTVRGWLHKEVKNXSPMSSLKUKKGFVLTHNSLDVYKSEKNAIKLGT 1440  
 Db 1382 SKGDTVEGGEFTVRGWLHKEVKNXSPMSSLKUKKGFVLTHNSLDVYKSEKNAIKLGT 1441  
 QY 1441 LVNLSICSVVPPEKFKETGVNWNVYGRKHCRILYTKLNAATRMSSYIQVVTPTKAP 1500  
 Db 1442 LVNLSICSVVPPEKFKETGVNWNVYGRKHCRILYTKLNAATRMSSYIQVVTPTKAP 1501  
 QY 1501 IDPTQQLIDIDKENCINSDVVEQIYKRNPLRYTHHPLHSPLPLPYGGINNLKDKG 1560  
 Db 1502 IDPTQQLIDIDKENCINSDVVEQIYKRNPLRYTHHPLHSPLPLPYGGINNLKDKG 1561  
 QY 1561 YTTLODEBAIKFNSLQOLBSMSDPPIIOGILOTGHDRLRLDELYCOLIKQTNKVPHNG 1620  
 Db 1562 YTTLODEBAIKFNSLQOLBSMSDPPIIOGILOTGHDRLRLDELYCOLIKQTNKVPHNG 1621  
 QY 1621 SVGNLVSWOILLTCLSCFLPSRGILKYLKFKHLKRIRBQFGTMEKVALFTYSLKKTTC 1680  
 Db 1622 SVGNLVSWOILLTCLSCFLPSRGILKYLKFKHLKRIRBQFGTMEKVALFTYSLKKTTC 1681  
 QY 1681 REFVPSRDEIEALIHROEMTSTVYCHGGSCKTITNSHTTAGVEVEKLIRGLAMEDSRNM 1740  
 Db 1682 REFVPSRDEIEALIHROEMTSTVYCHGGSCKTITNSHTTAGVEVEKLIRGLAMEDSRNM 1741  
 QY 1741 FALFENGHVDAKIEESTVADVLAKEKLAATSEVDLFWKCFPKLYCFGLDTDNVPKDS 1800  
 Db 1742 FALFENGHVDAKIEESTVADVLAKEKLAATSEVDLFWKCFPKLYCFGLDTDNVPKDS 1801  
 QY 1801 VEEAFMEQAEHNVIGHHHPAREBNQVLAALRLOLYLOGYTLHAALPRIEEVYSIORLK 1860  
 Db 1802 VEEAFMEQAEHNVIGHHHPAREBNQVLAALRLOLYLOGYTLHAALPRIEEVYSIORLK 1861  
 QY 1861 ARIISQSTKTTPCEPERLEKRTSPLEGLTRSPFGTSVVRQKVEEOMLDMWIEEVSAR 1920  
 Db 1862 ARIISQSTKTTPCEPERLEKRTSPLEGLTRSPFGTSVVRQKVEEOMLDMWIEEVSAR 1921  
 QY 1921 ASIIDKWRKFQGNNOEQAMAKYALIKEMPGYSTLFDEVCXKGEFPOELMIGVSADAYS 1980  
 Db 1922 ASIIDKWRKFQGNNOEQAMAKYALIKEMPGYSTLFDEVCXKGEFPOELMIGVSADAYS 1981  
 QY 1981 VYKRGGRPLREVOYEHILISFGAPLANTYKIVDERELLFETSEVVDVAKLMAIYSIMV 2040  
 Db 1982 VYKRGGRPLREVOYEHILISFGAPLANTYKIVDERELLFETSEVVDVAKLMAIYSIMV 2041

QY 2041 KRRYSTTRSSASSQSSR 2057  
 Db 2042 KRRYSTTRSSASSQSSR 2058  
 RESULT 3  
 AAE11891  
 ID AAE11891 standard; Protein; 2048 AA.  
 AAE11891;  
 18-DEC-2001 (first entry)  
 Angiogenesis associated human myosin X2 (hMX2) protein variant.  
 Angiogenesis associated protein; AAP; cytosolic; cardiac; gene therapy;  
 ophthalmological; vulnary; myocardial infarction; macular degeneration;  
 diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;  
 rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;  
 cancer; therapeutic; diagnostic; human; myosin X2; MX2; mutant; mutant;  
 variant.  
 Homo sapiens.  
 Synthetic.  
 MO200170808-A2.  
 27-SEP-2001.  
 22-MAR-2001; 2001WO-US09609.  
 22-MAR-2000; 2000US-191134P.  
 (CURA-) CURAGEN CORP.  
 (GENT) GENENTECH INC.  
 Rastelli LK, Gerritsen M;  
 WPI; 2001-60275/68.  
 DR N-PSDB; AAD19120.  
 Novel angiogenesis associated polypeptides and polynucleotides encoding  
 the polypeptides, useful for modulating angiogenesis and for treating  
 tumors and cancers -  
 Claim 1; Page 42-49; 159pp; English.  
 The invention relates to angiogenesis associated proteins (AAP) and their  
 corresponding cDNA molecules, which are useful for modulating  
 angiogenesis. AAP proteins and nucleic acids are useful for promoting  
 wound healing, for example after organ transplantation, and in the  
 treatment of tumours, myocardial infarction, cancers, diabetic  
 retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.  
 AAP proteins and DNA's are useful in potential prophylactic and  
 therapeutic applications implicated in a variety of disorders including  
 those related to angiogenesis, and also in diagnostic applications.  
 AAP cDNA is also useful in gene therapy. The invention also relates to  
 a method for screening a tissue sample for tumourigenic potential. AAP  
 proteins are used to screen drugs or compounds that modulate AAP activity  
 or expression as well as treating disorders characterised by insufficient  
 or excessive production of AAP or production of AAP forms that have  
 decreased or aberrant activity compared to the wild type protein, or  
 modulate biological function that involve AAP. The present sequence  
 is human myosin X2 (hMX2) variant protein which is an angiogenesis  
 associated protein (AAP) of the invention. hMX2 is upregulated in the in  
 vitro model of angiogenesis and is likely to be involved in transducing  
 signals from angiogenic factors, perhaps modulating the cytoskeleton.  
 SO Sequence 2048 AA;  
 Query Match 95.5%; Score 10247; DB 22; Length 2048;  
 Best Local Similarity 99.8%; Pred. No. 0;

Matches 1961; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	FCLOGRVWLRENGQHFPSTVNSCAGIYVFRDYGOVFPYKOSTITTHOKTAMHPTEER	60
Db	1	FCLOGRVWLRENGQHFPSTVNSCAGIYVFRDYGOVFPYKOSTITTHOKTAMHPTEER	60
Qy	61	GVDDMASTLTELHGGSIMYNLFQRYKKNQIWTYIGSILASVNPQIAGLYEPATMBOYR	120
Db	61	GVDDMASTLTELHGGSIMYNLFQRYKKNQIWTYIGSILASVNPQIAGLYEPATMBOYR	120
Qy	121	RHLGELPRPHI FALANECYRCMLKRNOCCLIKGEGAGITESTKILKFLSTISQOQLB	180
Db	121	RHLGELPRPHI FALANECYRCMLKRNOCCLIKGEGAGITESTKILKFLSTISQOQLB	180
Qy	181	LSLKEKTSCEVERAIILESSPIMEAFNAKTYVNNNSRFGIFVQANI CQKNGIOGRIYVC	240
Db	181	LSLKEKTSCEVERAIILESSPIMEAFNAKTYVNNNSRFGIFVQANI CQKNGIOGRIYVC	240
Qy	241	ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREFEYLSIPENYHYLANOGCVEDKTIID	300
Db	241	ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREFEYLSIPENYHYLANOGCVEDKTIID	300
Qy	301	QSSFRREVITAMDMVQMSKEEVRSRLAGILHGNIEPITTAGAQSFPKTALGRSAELL	360
Db	301	QSSFRREVITAMDMVQMSKEEVRSRLAGILHGNIEPITTAGAQSFPKTALGRSAELL	360
Qy	361	GLDPTQLTALTORSMPLRGEELITPLNQOAVDSRSLAMALYACCFEVIKIKINSRIK	420
Db	361	GLDPTQLTALTORSMPLRGEELITPLNQOAVDSRSLAMALYACCFEVIKIKINSRIK	420
Qy	421	GNEDFKSGILDI FGFEENFEVNHFEQFININYANEKLOEYFNKHI FSLQLEYSREGIYWE	480
Db	421	GNEDFKSGILDI FGFEENFEVNHFEQFININYANEKLOEYFNKHI FSLQLEYSREGIYWE	480
Qy	481	DIDMIDNGECLDIIEKKGLIALLINESHPQATDSTLLEKLSQHANNHFTYKPVAVN	540
Db	481	DIDMIDNGECLDIIEKKGLIALLINESHPQATDSTLLEKLSQHANNHFTYKPVAVN	540
Qy	541	NFGVKHAGEVOYDVAGILEKONDTFRDPLNLRSRPDIYDLPEHVSRRNOOTLKC	600
Db	541	NFGVKHAGEVOYDVAGILEKONDTFRDPLNLRSRPDIYDLPEHVSRRNOOTLKC	600
Qy	601	GSKHRRPTVVSQFVDSLHSLMATLSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Db	601	GSKHRRPTVVSQFVDSLHSLMATLSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Qy	661	MEETVARIKAGAVRRPFDQFYRYKYLKRNLLPEDVNGKCTSLIQLYDASNSEWOLGK	720
Db	661	MEETVARIKAGAVRRPFDQFYRYKYLKRNLLPEDVNGKCTSLIQLYDASNSEWOLGK	720
Qy	721	TKVFLRESLEOKLEKREEVSHAAVYIRAHVYGFARQYRVLYCWTIOKNYAPFL	780
Db	721	TKVFLRESLEOKLEKREEVSHAAVYIRAHVYGFARQYRVLYCWTIOKNYAPFL	780
Qy	781	RRRFLHLKKAAYIFOKOLRQIARVYRQLLAEKREOEKKQOESEKKKREBEERER	840
Db	781	RRRFLHLKKAAYIFOKOLRQIARVYRQLLAEKREOEKKQOESEKKKREBEERER	840
Qy	841	ERRREALRQOEETKQOELALOKSOKAEALTRELEKQENKQOYBELIRLEKEIEDIQ	900
Db	841	ERRREALRQOEETKQOELALOKSOKAEALTRELEKQENKQOYBELIRLEKEIEDIQ	900
Qy	901	RMKEQOELSTELASLOKLOERRDOELRLLEEAACRAQHFLESINDEIDECVNRTERSL	960
Db	901	RMKEQOELSTELASLOKLOERRDOELRLLEEAACRAQHFLESINDEIDECVNRTERSL	960
Qy	961	SGGSSEFSSELAESACEKPNFNSQPYPEEVEDEGEALDDAKDSPNSEHSHSDQRTS	1020
Db	961	SGGSSEFSSELAESACEKPNFNSQPYPEEVEDEGEALDDAKDSPNSEHSHSDQRTS	1020
Qy	1021	GRTSDSDSEEDPYNDTVVPTSPSADSTVLLAPSVQDSGLHNSSGSSTYCMPONAGD	1080
Db	1021	GRTSDSDSEEDPYNDTVVPTSPSADSTVLLAPSVQDSGLHNSSGSSTYCMPONAGD	1080

Qy	1081	LPSPDGDYDYDDYEDGALTSSSVTSNSYSGQMSPDYRCSVGTYNSSGAYRFSSEGA	1140
Db	1081	LPSPDGDYDYDDYEDGALTSSSVTSNSYSGQMSPDYRCSVGTYNSSGAYRFSSEGA	1140
Qy	1141	QSSFEDSEEDFSRFDTDDELRYRDSVYSCVTLTPYFPHSLYMKGLANSMKRRVCVLKD	1200
Db	1141	QSSFEDSEEDFSRFDTDDELRYRDSVYSCVTLTPYFPHSLYMKGLANSMKRRVCVLKD	1200
Qy	1201	ETFLMFRSKOZALKOGMLHKGGGSTLSRRWKKRWFLRQSKLAMPNDSEBKXGVN	1260
Db	1201	ETFLMFRSKOZALKOGMLHKGGGSTLSRRWKKRWFLRQSKLAMPNDSEBKXGVN	1260
Qy	1261	EVRTAKEIINDNTKNGIDIIIMADRTPHLIAESPDAQMFSLQVHASTQOEIOEMND	1320
Db	1261	EVRTAKEIINDNTKNGIDIIIMADRTPHLIAESPDAQMFSLQVHASTQOEIOEMND	1320
Qy	1321	EQANPONAVGTLVDGLIDVSCASDSPDRPNSFVITIANRVLHCNADTPEEMHWTLLQR	1380
Db	1321	EQANPONAVGTLVDGLIDVSCASDSPDRPNSFVITIANRVLHCNADTPEEMHWTLLQR	1380
Qy	1381	SKGDRVRGQEFIVRGMLHKEVKNSPKMSLKLKRWFLVLTNLSLDYKSSRKNLXICT	1440
Db	1381	SKGDRVRGQEFIVRGMLHKEVKNSPKMSLKLKRWFLVLTNLSLDYKSSRKNLXICT	1440
Qy	1441	LVNLNSLCSVPPDEKIFKETGYMNTVYGRKHCYRLYTLQLNEATRWSVIONVDTKAP	1500
Db	1441	LVNLNSLCSVPPDEKIFKETGYMNTVYGRKHCYRLYTLQLNEATRWSVIONVDTKAP	1500
Qy	1501	IDPTQQLIODIKENCLNSDVEQIYKRNPILRVTHHPHLSPLPLPYGDINLNLKDXG	1560
Db	1501	IDPTQQLIODIKENCLNSDVEQIYKRNPILRVTHHPHLSPLPLPYGDINLNLKDXG	1560
Qy	1561	YTTLDDEAKIKINSIQOLESMSDPIPIIQGLIQTHDLRPLDELYCOLIKQTNKVPHG	1620
Db	1561	YTTLDDEAKIKINSIQOLESMSDPIPIIQGLIQTHDLRPLDELYCOLIKQTNKVPHG	1620
Qy	1621	SVGNLYSMOILCTCLCTFLPSRGIKYLKFLHKLRIEOPPGTEMEKVALFTYESLKTTC	1680
Db	1621	SVGNLYSMOILCTCLCTFLPSRGIKYLKFLHKLRIEOPPGTEMEKVALFTYESLKTTC	1680
Qy	1681	REFVPSRDEIALIHRQEWSTFVYCHGGGCKITINSHTAGVYEKILRGILAMEDSRNM	1740
Db	1681	REFVPSRDEIALIHRQEWSTFVYCHGGGCKITINSHTAGVYEKILRGILAMEDSRNM	1740
Qy	1741	FALFEYNGHVDQAIISRIVYADVLAKEFKLATTSVGDLPKMFYKLYCFDLDNVPKOS	1800
Db	1741	FALFEYNGHVDQAIISRIVYADVLAKEFKLATTSVGDLPKMFYKLYCFDLDNVPKOS	1800
Qy	1801	VEFAFMFEQAHBAVYHGHHPAPEENLOVLAALRLQYLODYYTLHAIPLEEYSLQRLK	1860
Db	1801	VEFAFMFEQAHBAVYHGHHPAPEENLOVLAALRLQYLODYYTLHAIPLEEYSLQRLK	1860
Qy	1861	ARISQSTKTFPCERLEKRRISFLEGTLRSSFRTSVVRQYVEECMLDMMIKEEVSAR	1920
Db	1861	ARISQSTKTFPCERLEKRRISFLEGTLRSSFRTSVVRQYVEECMLDMMIKEEVSAR	1920
Qy	1921	ASIIDKMRKFOGMOBOAMAKYMALIKEMPGGSLTFDVECKEG 1964	
Db	1921	ASIIDKMRKFOGMOBOAMAKYMALIKEMPGGSLTFDVECKEG 1964	

RESULT 4  
 ID ABG10631 standard; Protein; 2408 AA.  
 AC ABG10631;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #10622.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;





QY 1878 KRTSFLBGLRLRSFRTGSVVRQVEEQMLDMWIKIEBVSASASIIDKMRKEQGNQOQ 1937  
DB 1574 KRTSFLBGLRLRSFRTGSVVRQVEEQMLDMWIKIEBVSASASIIDKMRKEQGNQOQ 1633  
QY 1938 AAKAKYALIKEMPGYSTLFDVECKGKGFPQELMLGVADANVSYYRGGRPLEVFQYH 1997  
DB 1634 AAKAKYALIKEMPGYSTLFDVECKGKGFPQELMLGVADANVSYYRGGRPLEVFQYH 1693  
QY 1998 ILSFGAPLANTYIVVDERELFETSEBVYVAKLMAKAYISMTYKXYSSTRASASGSSSR 2057  
DB 1694 ILSFGAPLANTYIVVDERELFETSEBVYVAKLMAKAYISMTYKXYSSTRASASGSSSR 1753

RESULT 5  
ABG10624  
ID ABG10624 standard; Protein; 1515 AA.  
XX  
XX ABG10624;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #10615.  
DE  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX MPI; 2001-639362/73.  
DR N-PSDB; AAS74811.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
XX Claim 20; SEQ ID No 40983; 103pp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX  
SQ Sequence 1515 AA;  
Query Match 63.1%; Score 6775.5; DB 22; Length 1515;  
Best Local Similarity 93.3%; Pred. No. 0;  
Matches 1342; Conservative 6; Mismatches 8; Indels 81; Gaps 5;  
QY 641 QKMPQPDQAVVLANOLARYSGMLETYRIRKAGYAVRPPQDFYKRYKYLNRNALPEDVNG 700  
DB 180 RMPMPQPDQAVVLANOLARYSGMLETYRIRKAGYAVRPPQDFYKRYKYLNRNALPEDVNG 226  
QY 701 KCTSLLOLYDASNSWQLGKTVPLRESLEQLEKREBEVSHAMVIRAHVGLFLARKQ 760  
DB 227 KCTSLLOLYDASNSWQLGKTVPLRESLEQLEKREBEVSHAMVIRAHVGLFLARKQ 286  
QY 761 YKRVLCVVTIOKNYRAFLRRRFLHLKKAALVPOKRGQIARVYQOLAEXREOEK 820  
DB 287 YKRVLCVVTIOKNYRAFLRRRFLHLKKAALVPOKRGQIARVYQOLAEXREOEK 346  
QY 821 KQEEEEKKKREBEEREREREREAELAQAQEEETRKOQLEALQKQKAEALTRLEKQ 880  
DB 347 KQEEEEKKKREBEEREREREREAELAQAQEEETRKOQLEALQKQKAEALTRLEKQ 406  
QY 881 KENKQVEETLRLEKEIEDLQRMKEQOELSTFASLQKLQERRDOELRLLEEACPAQEF 940  
DB 407 KENKQVEETLRLEKEIEDLQRMKEQOELSTFASLQKLQERRDOELRLLEEACPAQEF 466  
QY 941 LESLNPDEIDECVNIERSLSGSGSFSSGLASACEKPNFNSQPYPEEYDEGFELAD 1000  
DB 467 LESLNPDEIDECVNIERSLSGSGSFSSGLASACEKPNFNSQPYPEEYDEGFELAD 526  
QY 1001 DAFKDSPNPSEHGHDDQRTSGIRTSDDSEEDPYNNDTVPFSPADSTVLLAPVQDSG 1060  
DB 527 DAFKDSPNPSEHGHDDQRTSGIRTSDDSEEDPYNNDTVPFSPADSTVLLAPVQDSG 568  
QY 1061 SLHNSSGSESTYCMPQNAAGDLPSPDGDDYDDYDDEGATISGSSVTSNSYGSQMSPDY 1120  
DB 569 SLHNSSGSESTYCMPQNAAGDLPSPDGDDYDDYDDEGATISGSSVTSNSYGSQMSPDY 606  
QY 1121 RCVSGTYNSSGAYRFSRSGAQSFPEDSEDPSPRPTDDELRYRDSVYSCVTLVYFHSF 1180  
DB 607 RCVSGTYNSSGAYRFSRSGAQSFPEDSEDPSPRPTDDELRYRDSVYSCVTLVYFHSF 666  
QY 1181 LYMKGLMNSWGRKRCVLDKDETFPLFRSKQELKKGWMLHKKGSGSTLSRRMKKRWFL 1240  
DB 667 LYMKGLMNSWGRKRCVLDKDETFPLFRSKQELKKGWMLHKKGSGSTLSRRMKKRWFL 726  
QY 1241 ROSKLMYFENDSEBKLGTVFVRTAKEIIDNTTKENGIDIIWADRTPHLIASEPDAQW 1300  
DB 727 ROSKLMYFENDSEBKLGTVFVRTAKEIIDNTTKENGIDIIWADRTPHLIASEPDAQW 786  
QY 1301 ESYLSQVHASTQOEIOEHMDEQANPQNAVGLTDLGLISVCSDSFDRPNSVITIANRV 1360  
DB 787 ESYLSQVHASTQOEIOEHMDEQANPQNAVGLTDLGLISVCSDSFDRPNSVITIANRV 846  
QY 1361 LHCNADTPBEMHMTTLQORSKGDTRVGEQETVGMHLKEVKNSPKMSLTLKRWFL 1420  
DB 847 LHCNADTPBEMHMTTLQORSKGDTRVGEQETVGMHLKEVKNSPKMSLTLKRWFL 906  
QY 1421 THNSLDYKSSBKNAIKLGTVLNLSGVPDEKIFKETGYMNTVYGRKCYRLYTXL 1480  
DB 907 THNSLDYKSSBKNAIKLGTVLNLSGVPDEKIFKETGYMNTVYGRKCYRLYTXL 966  
QY 1481 LNEARWSSVIONVDTAPITPTPQOILQDIKENCASDVVEQIYKRNPIRLYTHHPH 1540  
DB 967 LNEARWSSVIONVDTAPITPTPQOILQDIKENCASDVVEQIYKRNPIRLYTHHPH 1026  
QY 1541 SPLPLPAGDILANTLKKGYTTTLODEAIKFNLSQQLESMSDPIITIGILOTGHDLRP 1600  
DB 1027 SPLPLPAGDILANTLKKGYTTTLODEAIKFNLSQQLESMSDPIITIGILOTGHDLRP 1086  
QY 1601 LBDLVCQILKQTNKVPFPGSVGNLYSQOILTGLSCTPLPSRGIKLYLPHLKRIREOPF 1660  
DB 1086 LBDLVCQILKQTNKVPFPGSVGNLYSQOILTGLSCTPLPSRGIKLYLPHLKRIREOPF 1660

Db 1087 LRDELVYCOLIKQNNKVPHPGSGVNLVSMQILTCLSCFLPBGILKYLKFKHLKRIEOPF 1146  
Qy 1661 GTEMEXKALFTYESLKTCKREFVPSRDELEALIHREMTSTVYCHGGGCKITINSHTT 1120  
Db 1147 GTEMEXKALFTYESLKTCKREFVPSRDELEALIHREMTSTVYCHGGGCKITINSHTT 1206  
Qy 1721 AGEVVEKILIRGLAMEDSRNMFALFEYNGVDKAIESRTVADVLAKEKLAATSEVGDLP 1780  
Db 1207 AGEVVEKILIRGLAMEDSRNMFALFEYNGVDKAIESRTVADVLAKEKLAATSEVGDLP 1261  
Qy 1781 WKFEYKLYCFLDNDVNEKDSVEFAFMEOHNEAVIHGHHPAPERNLOVLAALQYLQGD 1840  
Db 1262 -----LTVSF-----QHAEVVIHGHHPAPERNLOVLAALQYLQGD 1298  
Qy 1841 YTLHAAIPLEEVYSIORLKARISQSTKFTPCERLEKRTSFLEGLTRSPRTGSVVRQ 1900  
Db 1299 YTLHAAIPLEEVYSIORLKARISQSTKFTPCERLEKRTSFLEGLTRSPRTGSVVRQ 1358  
Qy 1901 KVEEQLMDWMIKEEVSSARASIIDKWRKFGQNNQOQAKYMALIKEMPGYSTLPDVE 1960  
Db 1359 KVEEQLMDWMIKEEVSSARASIIDKWRKFGQNNQOQAKYMALIKEMPGYSTLPDVE 1418  
Qy 1961 CKRGFPQELMLGVSDAVSVYKRGGRPLEVFOYEHILSFQAPLANTYKIVDERELLP 2020  
Db 1419 CKRGFPQELMLGVSDAVSVYKRGGRPLEVFOYEHILSFQAPLANTYKIVDERELLP 1478  
Qy 2021 ETESEVVDVAKLMKAYISMIYKKRYSTTRSSASQSSSR 2057  
Db 1479 ETESEVVDVAKLMKAYISMIYKKRYSTTRSSASQSSSR 1515

RESULT 6  
AAU17082  
ID AAU17082 standard; Protein: 1031 AA.

XX AAU17082;

DT 07-NOV-2001 (first entry)

XX Novel signal transduction pathway protein, Seq ID 647.

XX Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;  
XX antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
XX Immune system disorder; rheumatoid arthritis; inflammatory condition;  
XX organ transplant rejection; infection; hepatitis C; blood disorder;  
XX sickle cell anemia; hyperproliferative disorder; Gaucher's disease;  
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
XX chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
XX cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
XX reproductive system; gastrointestinal; liver disorder; AIDS;  
XX acquired immune deficiency syndrome.

OS Homo sapiens.

XX MO200154733-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001MO-US01312.

XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0180664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225577.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.



Db	786	SAVIGHHNPADENQVLAALRLQYLQSGDYTLHMAIPLEBEVYLSQRKANIISOSTKFT	845
Qy	1872	PCERLEKRTSFLEBTLRFSRFTGVSVPQKBEEDMLMWIKEVSSARASIIDKRRFQ	19311
Db	846	PCERLEKRTSFLEBTLRFSRFTGVSVPQKBEEDMLMWIKEVSSARASIIDKRRFQ	905
Qy	1932	GMNDOAAMAKMALIKEMPGYSTLFDVECKEGGFPQELMTGVSADAVSVKRGGRPLE	19921
Db	906	GMNDOAAMAKMALIKEMPGYSTLFDVECKEGGFPQELMTGVSADAVSVKRGGRPLE	965
Qy	1992	VFOYEHILSPGAPLANTYKIIVDERELLFETISEVVDVAKMLKAYISMTVKRYSSTRSAS	20511
Db	966	VFOYEHILSPGAPLANTYKIIVDERELLFETISEVVDVAKMLKAYISMTVKRYSSTRSAS	10251
Qy	2052	SOQSSR 2057	
Db	1026	SOQSSR 1031	

XX	RESULT 7		
XX	AA093762		
XX	ID	AA093762 standard; Protein; 984 AA.	
XX	AA093762;		
XX	AC		
XX	06-NOV-2001	(first entry)	
XX	DE	Human polypeptide, SEQ ID NO: 3757.	
XX	KW	Human, full length cDNA; cDNA synthesis; oligo-capping.	
XX	OS	Homo sapiens.	
XX	PN	EP130094-A2.	
XX	PD	05-SEP-2001.	
XX	PF	07-JUL-2000; 2000EP-0114069.	
XX	PR	08-JUL-1999; 99JP-0194486.	
XX	PR	11-JAN-2000; 2000JP-0118774.	
XX	PR	02-MAY-2000; 2000JP-0183765.	
XX	PA	(HELI-) HELIX RES INST.	
XX	PI	Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawat Y;	
XX	PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	DR	WPI: 2001-524255/58.	
XX	DR	N-PSDB: AAK94716.	
XX	PT	830 Primers useful for synthesizing full length cDNA clones and their	
XX	PT	use in genetic manipulation -	
XX	PS	Claim 8; SEQ ID NO 3757; 1380bp + sequence listing; English.	
XX	CC	The invention relates to primers for synthesizing full length cDNA	
XX	CC	clones. 830 cDNA molecules encoding a human protein have been	
XX	CC	isolated and nucleotide sequences of 5' and 3'-ends of the cDNA	
XX	CC	molecules have been determined. Primers for synthesizing the full length	
XX	CC	cDNA are useful for clarifying the function of the protein encoded by	
XX	CC	the cDNA. The full length clones were obtained by construction of full	
XX	CC	length enriched cDNA libraries that were synthesized by the oligo-capping	
XX	CC	method. The primers enable the production of the full length cDNA easily	
XX	CC	without any special methods. The present sequence is a polypeptide	
XX	CC	encoded by a full length human cDNA of the invention.	
XX	CC	Note: The sequence data for this patent did not form part of the printed	
XX	CC	specification, but was obtained in CD-ROM format directly from EPO.	
XX	SQ	Sequence 984 AA;	
XX	Query Match	48.1%; Score 5166;	DB 22: Length 984;
XX	Best Local Similarity	99.6%; Ptd. No. 0;	

Matches 980; Conservative 1; Mismatches 3; Indels 0; Gaps 0.

QY	1074	MPONAGDLPSPDGDYDYODDYEDQAIIRSGSVTSNSVGSQMSDPYCSVCTYNSGAY	113
Db	1	MPONAGDLPSPDGDYDYODDYEDCATIRSGSVTFSNSVGSQMSDPYCSVCTYNSGAY	60
QY	1134	RFSBGAQSSPDESEDEDFSRFDTDDELSTYRDSVYSCVTLPEYFHSFLYMGGLNMSWR	119
Db	61	RFSBGAQSSPDESEDEDFSRFDTDDELSTYRDSVYSCVTLPEYFHSFLYMGGLNMSWR	120
QY	1194	RMVCLKDTFLMFRKQKALQGMHLKKGGSSTLSRNNMKRMVLAOSKLMYFENSE	125
Db	121	RMVCLKDTFLMFRKQKALQGMHLKKGGSSTLSRNNMKRMVLAOSKLMYFENSE	180
QY	1254	EKLKGTVEVTRAKELIDNTTKENGIDIMADRTFHLIAESPEDASQWFSVLSQVASTDQ	131
Db	181	EKLKGTVEVTRAKELIDNTTKENGIDIMADRTFHLIAESPEDASQWFSVLSQVASTDQ	240
QY	1314	EIQEMHDEQANPONAUGTLVDGLIDSVCASSDPDRPNSFVITITANRVLHCNADTEBEMH	137
Db	241	EIQEMHDEQANPONAUGTLVDGLIDSVCASSDPDRPNSFVITITANRVLHCNADTEBEMH	300
QY	1374	WITLLQBSKGTREVGQEFIVRGMLHKEVNSPKMSSLKLKRMVVLHNSLIDYKSSSK	143
Db	301	WITLLQBSKGTREVGQEFIVRGMLHKEVNSPKMSSLKLKRMVVLHNSLIDYKSSSK	360
QY	1434	NALKLGTVLVNSLCSVVPDEKIFKETGYMVTYVGRGHCRYLTYKLINATRMSSVJON	149
Db	361	NALKLGTVLVNSLCSVVPDEKIFKETGYMVTYVGRGHCRYLTYKLINATRMSSVJON	420
QY	1494	VTDTKAPIDTPTQOOLIDIKENCCLNSDVEQIYKKNPLIRYTHHPLHSLPLPYGINDL	155
Db	421	VTDTKAPIDTPTQOOLIDIKENCCLNSDVEQIYKKNPLIRYTHHPLHSLPLPYGINDL	480
QY	1554	NLLMDKGYTTLQDEAKIKFNSLQOLUESMDPIPIIOGLQTHGHDRLPRDELXYCOLIKOT	161
Db	481	NLLMDKGYTTLQDEAKIKFNSLQOLUESMDPIPIIOGLQTHGHDRLPRDELXYCOLIKOT	540
QY	1614	NKVPHPGSVGNLYSMQOILTCDSCTPLPSRGILKYLKFKHLIREOPGTEMEKVALFTYE	167
Db	541	NKVPHPGSVGNLYSMQOILTCDSCTPLPSRGILKYLKFKHLIREOPGTEMEKVALFTYE	600
QY	1674	SLKTKTCHEPSPSRDEIHALIHRQEMSTYVCHGGGSCKITINSHTTAGEVEVEKILRGIA	173
Db	601	SLKTKTCHEPSPSRDEIHALIHRQEMSTYVCHGGGSCKITINSHTTAGEVEVEKILRGIA	660
QY	1734	MEDSRNMFALPEYNGHVDAKIESRIVADVLAKEFKLATSEVGDLPMKFYKLCFLDPT	179
Db	661	MEDSRNMFALPEYNGHVDAKIESRIVADVLAKEFKLATSEVGDLPMKFYKLCFLDPT	720
QY	1794	DNVPKDSVEFAFMFEQAHNAVIGHHNPABEENLOVALALRIQYLOGDYTLHAALPBLEEV	185
Db	721	DNVPKDSVEFAFMFEQAHNAVIGHHNPABEENLOVALALRIQYLOGDYTLHAALPBLEEV	780
QY	1854	YSLORLKARIQOSTTFTPCERLEKRRISFLEGTLRBSFRFGSVVRQVYEEQMDMWIK	191
Db	781	YSLORLKARIQOSTTFTPCERLEKRRISFLEGTLRBSFRFGSVVRQVYEEQMDMWIK	840
QY	1914	BEVSARASIIDKMRKFOGMNOQOAMAKYMLIKEMPGYSTLPFVECKEGGFPOBLWLG	197
Db	841	BEVSARASIIDKMRKFOGMNOQOAMAKYMLIKEMPGYSTLPFVECKEGGFPOBLWLG	900
QY	1974	VSADAVSYKRGEGRPLEVFQYEHILSLFSGAPLANTYKIVDERELLFETSEVVDAKLMM	203
Db	901	VSADAVSYKRGEGRPLEVFQYEHILSLFSGAPLANTYKIVDERELLFETSEVVDAKLMM	960
QY	2034	AYISMIYVKRSTTRRSASSQSSSR 2057	
Db	961	AYISMIYVKRSTTRRSASSQSSSR 984	

RESULT 8  
ABG25601

RESULT 8  
ABG25601

ID ABG25601 standard; Protein; 1442 AA.  
XX  
AC ABG25601;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #25592.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-P8DB; AAS89788.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 55960; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensic, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1442 AA;  
XX  
Query Match 48.0%; Score 5146.5; DB 22; Length 1442;  
Best local similarity 83.7%; Pred. No. 0;  
Matches 1029; Conservative 31; Mismatches 82; Indels 87; Gaps 14;  
QY 555 VAGILEKNDOTFFDDLLN-----LIRESRFDFTYLFHVSRRNQDTLKCG-----601  
DB 214 IPELOAE-----FREDALNMGDEKIFKTELVNDMKINGRIRERASSTCEGLRIIYQ 269  
QY 602 -----SKIRRPVVSQ-----FKVDLSLSIMATLSSNPFVRCK---637  
DB 270 VSFHSSVHLFTNVGQVRRRIACERACMEYRSASLILNSLNSVEG--VYREKEITA 326  
QY 638 -----PNM-----QKMPDQFOAVVNLQRLRYSGILETVIRIKAGYAVRRPFQDFY 682

DB 327 AEKTTNGNDLPDPDRCRMPDPQFOAVVNLQRLRYSGILETVIRIKAG-----375  
QY 683 KRYKLMENIALPEVNRGKCTSLQLYASNSEWOLGKTVLRSLRQKLEKREBEVS 742  
DB 376 --YKVLNRLALPEVNRGKCTSLQLYASNSEWOLGKTVLRSLRQKLEKREBEVS 433  
QY 743 HAAWVIRAHVIGFLARKQYRKVLCVVIIOKNYRAFLRRRPLHLKKAIVFOQLRGCI 802  
DB 434 HAAWVIRAHVIGFLARKQYRKVLCVVIIOKNYRAFLRRRPLHLKKAIVFOQLRGCI 493  
QY 803 ARRVYRQLIAEKREDEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 862  
DB 494 ARRVYRQLIAEKREDEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 553  
QY 863 ALQSKQKAEILFRELEKKNKKNQVEILRLKEIETDLORMKEQOELSLTEASLOQER 922  
DB 554 ALQSKQKAEILFRELEKKNKKNQVEILRLKEIETDLORMKEQOELSLTEASLOQER 613  
QY 923 DQELRLLEEACRAAQEFLESINPDEIDECVANIERSLSGSEFSESLAESACEKPNFN 982  
DB 614 DQELRLLEEACRAAQEFLESINPDEIDECVANIERSLSVSEFSESLAESACEKPNFN 673  
QY 983 FSQPYP---EEEVDEGFADDDAFKDSPNBSEHSGSDQRTSGIRTSDDSSSEDPYKNDT 1038  
DB 674 LSKRPPPKRQEVDEGFADDDAFKDSPNBSEHSGSDQRTSGIRTSDDSSSEDPYKNDT 733  
QY 1039 VVPTSPADSTYLLAPSVQDSSSLNSSSGESTYCPQNAQGLPEPDGDYDIDODDYEDG 1098  
DB 734 VVPTSPADSTYLLAPSVQDSSSLNSSSGESTYCPQNAQGLPEPDGDYDIDODDYEDG 793  
QY 1099 AITSSSVTFPSNSYSGOWSPDYRCVGTYNSSGAYRFSSEGAQSFEDSEEDPDRPPTD 1158  
DB 794 AITSSSVTFPSNSYSGOWSPDYRCVGTYNSSGAYRFSSEGAQSFEDSEEDPDRPPTD 853  
QY 1159 DELSYRDSVYSCVTLPYFHSFLYMKGGLMNSWKRMCVLDQETFLMFRSKQALKQWL 1218  
DB 854 DELSYRDSVYSCVTLPYFHSFLYMKGGLMNSWKRMCVLDQETFLMFRSKQALKQWL 913  
QY 1219 HKKGGGSSLSLRNKKKFWLROSKLMYFENDSEKIKGVYVYTAKEIIDNTTKENCI 1278  
DB 914 HKKGGGSSLSLRNKKKFWLROSKLMYFENDSEKIKGVYVYTAKEIIDNTTKENCI 973  
QY 1279 DIIMADRFPHLAEPEPDASQFVLSQVHASTDDEIOEMDEQANPQNAVGTLLDVGILD 1338  
DB 974 DIIMADRFPHLAEPEPDASQFVLSQVHASTDDEIOEMDEQANPQNAVGTLLDVGILD 1033  
QY 1339 SVCASDSPRPNFVIITANRYLHCNADTPEEMHMTLL--QRSGDTRVEGOEPIVKGW 1397  
DB 1034 SVMASDSPRPNFVIITANRYLHCNADTPEEMHMTLLPEVAKGTPEMRARNSSXEDG 1093  
QY 1398 LKKEVKNBPKMSLTLKKRWFLL--TNSLDIYSSSEKNAKL--GTLVNSLCSVYP 1451  
DB 1094 CTKRRTYORCLHLTKKRWPFVLPQFPGLQOEF--SEPPKJGCPPTPTASAL--VVP 1149  
QY 1452 PDEKIFKEGYNNVYVGRKCYRLYTKLNBATWSSVIOQVNTDTPKAPIDPTQQLD 1511  
DB 1150 PDEKIFKEGYNNVYVGRKCYRLYTKLNBATWSSVIOQVNTDTPKAPIDPTQQLD 1209  
QY 1512 IKENCLNSDVYEQIYKRNPIRLYTHPLHSPPLPYGDINMLNLKDGYYTLQDEAKI 1571  
DB 1210 IKENCLNSDVYEQIYKRNPIRLYTHPLHSPPLPYGDINMLNLKDGYYTLQDEAKI 1269  
QY 1572 FNSLOQLESMSDPIPIIGILOTGDLARPLADELYCOLIKQTNKYPHGSVGNLWSQTL 1631  
DB 1270 FNSLOQLESMSDPIPIIGILOTGDLARPLADELYCOLIKQTNKYPHGSVGNLWSQTL 1329  
QY 1632 TGLSCTPLPSRGILKYLKFLKRIEOPPGTEMEKYALFTYESLKYTCRBPVPRDSE 1691  
DB 1330 TGLSCTPLPSRGILKYLKFLKRIEOPPGTEMEKYALFTYESLKYTCRBPVPRDSE 1389  
QY 1692 ALIHRQEMTSTVYCHGGSCKITINSHTT 1720

DB	1390	ALIHQEWSTVYCHGGGSCKITLNSHTT	1418
RESULT 9			
AAAB42020			
ID	AAAB42020	standard; Protein; 869 AA.	
XX			
AC	AAAB42020;		
XX			
DT	08-FEB-2001	(first entry)	
DE	Human ORFX ORF1784	polypeptide sequence SEQ ID NO:3568.	
XX			
KM	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnerability; antiparasitic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; anticharilic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal hemoglobinuria; burn; wound; thrombosis; cartilage damage; antiinflammatory disease; coagulation; chromosome; contraceptive.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200058473-A2.		
PD	05-OCT-2000.		
XX			
PF	31-MAR-2000; 2000WO-US08621.		
XX			
PR	31-MAR-1999; 99US-0127607.		
PR	02-APR-1999; 99US-0127636.		
PR	05-APR-1999; 99US-0127728.		
PR	30-MAR-2000; 2000US-0540763.		
XX			
PA	(CUBRA-) CUBAGEN CORP.		
XX			
P1	Shimkrets RA, Leach M;		
DR	WPI; 2000-602362/57.		
DR	N-PSDB; AAC76229.		
XX			
PT	Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -		
PS	Claim 11; Page 2727-2729; 5507pp; English.		
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnerability; antiparasitic; antiparkinsonian; nootropic; neuroprotective; immunostimulant; anticonvulsant; anticharilic; immunosuppressant; osteopathic; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; severe combined immunodeficiency (SCID); AIDS; viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,		

[illegible]







PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0244826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249247.  
PR 17-NOV-2000; 2000US-0249249.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251899.  
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465566/50.  
DR N-PSDB; AAS40999.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neutral, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases -  
XX  
PS Claim 11; SEQ ID NO 1125; 1180bp; English.  
XX  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 402 AA;  
  
Query Match 18.1%; Score 1947; DB 22; Length 402;  
Best Local Similarity 98.2%; Pred. No. 2.9e-127;  
Matches 378; Conservative 2; Mismatches 3; Indels 2; Gaps 2;  
  
QY 236 RIVDCILSNRVRNQNQNGERNYHIFVALLAGLEHEEREFPYLTSPENYHYNOSGCVED 295  
DB 20 RIVDYLL-EKRRVAVQNGERNYHIFVALLAGLEHEEREFPYLTSPENYHYNOSGCVED 78  
QY 236 KTISDQBSFREYITAMDVQPSKEEVRVSRLLAGILHGNIEFTTAGAQSFTALGR 355  
DB 79 KTISDQBSFREYITAMDVQPSKEEVRVSRLLAGILHGNIEFTTAGAQSFTALGR 138  
QY 356 SAEILGLDPTQTLTALTRSFPLRGEELTPLNVQAVDSRSLMALIYACCFEYVIXI 415  
DB 139 SAEILGLDPTQTLTALTRSFPLRGEELTPLNVQAVDSRSLMALIYACCFEYVIXI 198  
QY 416 NSRIKGNEDPFSIGLIDFGFENFEVNFEDQINIVAEKLOEYFNKHFISLEQYSRE 475  
DB 199 NSRIKGNEDPFSIGLIDFGFENFEVNFEDQINIVAEKLOEYFNKHFISLEQYSRE 258  
QY 476 GLVWEDIDWINGECLDLEIEKKLGLLALINESHPQATDSTLLEKLSQAHNNHFFYVYK 535  
DB 259 GLVWEDIDWINGECLDLEIEKKLGLLALINESHPQATDSTLLEKLSQAHNNHFFYVYK 318  
QY 536 RYAVNPFVGVKHYAGVQVDVKGILEKNDTRDDLLNLRSRPFPIYDLPEHVSRRNQ 595  
DB 319 RYAVNPFVGVKHYAGVQVDVKGILEKNDTRDDLLNLRSRPFPIYDLPEHVSRRNQ 378  
QY 596 DTLKGSKRRRPTVSSQFVSLNS 620  
DB 379 DTLKGSKRRRPTVSSQFVSLNS 402  
  
RESULT 13



QY 1245 ----LMPFENDSEKLTGTEVTRAKEIIDNT-----KENGIDIIADRTPHLIAES 1293  
 DB 1249 PIMLPITFMDGNTKTLIADS--ATTARELCNQLSDKISLKQDGFSLYIA--LFDKVSLSL 1304  
 QY 1294 PEDDAQWFEVLSQVNA-STDOELQE-----MHDSQANP---QNAV 1329  
 DB 1305 GSGGDHWDIAISQCEQYAKEQGAQENNAWRLEPRKEIAPMHEPTDDVANLIYQVAV 1364  
 QY 1330 GTLDVGLDVSQASDPDRPNSFVITANRVLIHCNADTEEMHHMTTLQDSKGDTRVAG 1389  
 DB 1365 RQVKG--EYRC-----DKEDDLAMIAAQYF--IEXTSMSERLFTLLPNITPDCLSG 1416  
 QY 1390 QEFIVRGW-----LHKEVKNSPKXSLKLK-----RMFVLTNLSLDYKSSERN 1434  
 DB 1417 VKAIERNALVIAQVAKSYVVDKIAPIKEDIYVAKYKWPPLFSFYEAVYNSGN 1476  
 QY 1435 ALKLTGLVINSLS--VPPDEKIFKETGYWNTVYGRHCHRLYKLIANEAT----- 1485  
 DB 1477 LFKNDVILAVNWTGVYVDDQVLLLESPPEITAVSSQTKNKFVQTSLSLTVRGEPT 1536  
 QY 1486 -----RMSSVIONVTDTPAKIDPT-----QQLIDIKENCIN 1518  
 DB 1537 FQSPNAEDRDVLVYVFLDGLKKRSKYVIALODYRABSDTSTLSPFKDILLDESCGE 1596  
 QY 1519 SDV-----VEQIYKRNPIRLYTHPLHSPLLPLPYGDI-----NLNLK 1557  
 DB 1597 SYLNNQMCICGRCDRSGERDPAETVYVL--PTLSKRPQDILAFNIEBAHHGRRLSMAS 1654  
 QY 1558 DKGYT-----TLQDEAIKINSLOQLSMS-----DPI--PIIO 1589  
 DB 1655 NGAVEPRDRPRLMEYALDHPR--LPPKRTMSKTLTSLSKRSBELWRSRDIKAPYLAK 1713  
 QY 1590 -----GILQTHDL--RP-----LRDELVCQLK 1611  
 DB 1714 LOSKEPFAEACPAAPAILKTMGDLPKPRMGNETTDIIFQGPLKHEILREBITYQLK 1773  
 QY 1612 QTNKVPHPGSGVNLYSWQILCTLSCTPLRSGLIKYKFKHRIROFPGTEMEKALPT 1671  
 DB 1774 QL--TDNRMRMBEERGMELMWLATGLFACSGGLKELFLKTRRRPIQDSMHR----- 1826  
 QY 1672 YSLSKT---KREFPVSRDEIEBALIHR--QEMTSTYVCHGGGSCKTINSHTTAGEVAK 1727  
 DB 1827 ---LQKTIHGORKYPHQVEVEALQKTOIFHKYFPPDDTDEAFEVDSSTRADFCNN 1883  
 QY 1728 LIRGLAMEDSRMFPAL-----FEVNGHVDAIESRTVAVAVLAKFEKL 1770  
 DB 1884 ISORLSRTSEG--FSLFVKIADKVISVPRGDFPDRHLTWIKKARPIRD----- 1934  
 QY 1771 AATSEVGDLPKRFYFLYCF--LDTDNV--KD--SVEFAFMFEQAEHAVIGHHAPAEENTL 1826  
 DB 1935 -----GANP--QFTYGVFFMKKLMNTVPGKORNDLIETHYQELPKLGRYHKCSREBA 1987  
 QY 1827 QVLAALRILOYLOGDYTLHAIRPLEEVYSLQRIKARIESGTTFTPCERLEKRTSFELEG 1886  
 DB 1988 AKLAALVFVRPGE-----NKOELQA--IFQMLRELIIPSDIMKIQSTS----- 2028  
 QY 1887 TLRBSRRGSVVRQKVEEOMLDMWIKESVSSARBSIIIMKMKFQOMNOEQAMAKYMALI 1946  
 DB 2029 -----EW-----KRISIVASYNODGMSDQKVAALVKV 2057  
 QY 1947 KEMPGVSTLFDV--ECKEGGFPOELMWIGSADAVSYKSGEGRPLEVFQYEHILSPGADL 2005  
 DB 2058 YMPFPGSAFPEVKQTEBNPEMLLIANKIGVSLIHHVTQDILVTHFTTISNMSS-- 2115  
 QY 2006 ANTY-----KIYVDERELLFETSEVVVDVAKLMKAYSNIV 2040  
 DB 2116 GNTYFMTIGNLVRGSKLCESTISLGYKMDLLTSYISLML 2155

AC AAU17508;  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Novel signal transduction pathway protein, Seq ID 1073.  
 XX  
 KW Neutropenic; cytostatic; dermatological; immunosuppressive; tumour;  
 KW antiinflammatory; anti-ILV; antibacterial; antiinflammatory; cancer;  
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
 KW organ transplant rejection; infection; hepatitis C; blood disorder;  
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 KW reproductive system; gastrointestinal; liver disorder; AIDS;  
 KW acquired immune deficiency syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200154733-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01312.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225477.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 14-AUG-2000; 2000US-0225799.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0228927.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.

RESULT 14  
 AAU17508  
 ID AAU17508 standard; Protein; 348 AA.  
 XX

PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0233223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250191.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-465460/50.  
DR N-PSDB; Aas27425.  
XX  
XX Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT diagnosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders -  
XX  
XX Claim 1, SEQ ID No 1073; 880pp; English.  
XX  
XX The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II), (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
CC Addison's disease), reproductive system disorders, gastrointestinal  
CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
CC as stimulators of B-cell responsiveness to pathogens, activators of  
CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency  
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
CC pathway protein, amino acid sequences of the invention.  
XX  
Query Match 16.7%; Score 1789; DB 22; Length 348;  
Best Local Similarity 99.1%; Pred. No. 2,5e-116;  
Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1033 DPYNDTVTPSPASDSTVTLAPSVQDSSGLHNSGSGSTYCPMPNAGDLPSPDGDYD 1091  
DB 6 DAWNDTVTPSPASDSTVTLAPSVQDSSGLHNSGSGSTYCPMPNAGDLPSPDGDYD 65  
QY 1092 QDDYEDGAIITSGSVTFNSYSGQWSPYRCVGTYNSSGAYRFSSEGNQSSFFEDSEDF 1151  
DB 66 QDDYEDGAIITSGSVTFNSYSGQWSPYRCVGTYNSSGAYRFSSEGNQSSFFEDSEDF 125  
QY 1152 DSRPDTDELSTYRDSVYSCVTLFYHSLFYKGLNWSKRRKCVLKDFTLFRSKOE 1211  
DB 126 DSRPDTDELSTYRDSVYSCVTLFYHSLFYKGLNWSKRRKCVLKDFTLFRSKOE 185  
QY 1212 ALKQGLMHLKGGSGSTLSRRNKKRPFVLRQSKLMYFENDSEKIKGTVEVTAKIIDN 1271  
DB 186 ALKQGLMHLKGGSGSTLSRRNKKRPFVLRQSKLMYFENDSEKIKGTVEVTAKIIDN 245

QY 1272 TTKENGIDIMADRTFHLIAESPEDASQMFVSLSOVHASTDIOEIOMEDEQANPONAVGT 1331  
DB 246 TTKENGIDIMADRTFHLIAESPEDASQMFVSLSOVHASTDIOEIOMEDEQANPONAVGT 305  
QY 1332 LDVGLIDSCASDSDPDRPNSFVITANRVLHCNADTPE 1369  
DB 306 LDVGLIDSCASDSDPDRPNSFVITANRVLHCNADTPE 343

RESULT 15  
ABB62828  
ID ABB62828 standard; Protein; 2129 AA.  
AC ABB62828;  
DT 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 11276.  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
OS Drosophila melanogaster.  
PN W0200171042-A2.  
PD 27-SEP-2001.  
PP 23-MAR-2001; 2001WO-US09231.  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-061415O.  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;  
DR MPI, 2001-656860/75.  
DR N-PSDB; ABL06931.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
PS Disclosure: SEQ ID NO 15276; 21bp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB12072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2129 AA;  
Query Match 15.9%; Score 1705; DB 22; Length 2129;  
Best Local Similarity 25.7%; Pred. No. 3.1e-109;  
Matches 598; Conservative 392; Mismatches 828; Indels 508; Gaps 84;

QY 3 LOGTAVMLRENGQHPSTNSCAE-----GIVFRFDYGVFLT---YQSTITTHOKV-T 52  
DB 13 ICGEYVWVWPQ-----NTTSEFVAVPGARIVREKQTQLVCDNRKQFWMPADVLK 64  
QY 53 AAHPTEEGVDDMASLTSLHGSIMYNLFQYKRNQIMTYGISILASVNPYQPIAGLYRP 112  
DB 65 AAHTISGEVDDMTLGDLOETILNLFQNRKAKOLITTYTSGMLVAINPYQ-ILPIYTN 123  
QY 113 ATMEQYSRRHLGELPPIFAIANECYRCIMKGRHNDCLIKESGAGKTESKILIKFLS 172

DB 124 REIQLYRKNLSLAELPPHIFALISDNFQRLQRIKENQCVVISGSESAGKTESKILIQYLA 183  
QY 173 VISOQSLELSLKERTSCVERAIILESSPIIMEAFGNKATVYNNNSNFFGKFOVNICQKNGI 232  
DB 184 AIS-----GKNSWIEQOIIENAPIMEAFGNKATVNDNSNFFGKITEIRFPQGI 234  
QY 233 QGGRIVDILSQNRNVKQNPGEKNYHIFVALAGLEHEEREEFYLL--STPENHYLNQS 290  
DB 235 QGARIQOYL--EKSLIVFQSRDERNYHIFYCLAGLSIAERRLQIQSGSPQYHYLNQ 293  
QY 291 GCVEDKTSIDQSEFREVITAMDVMOFSKEEYREVSRLAGILHGNIEFITA----GGA 345  
DB 294 GCFTLPGRGDAKDFADIRAAKVLSPKPEEWSILSLAAILHGLNLEFTEVANLATA 353  
QY 346 QVSEFTALGRSAELGLDPTQTLDTLQSRSMFLRGEELITPLNQVADSRLMALAYA 405  
DB 354 ETDTPNLRQVQAQLGIPISALNALTORITFVGEHVITSLSKMALGRDAFYKSLYD 413  
QY 406 CCFEWWIKINSRIKGNED--FKSIGIIDIFGFENFEVNHFEQFNINYNANEKLQAEYFNH 463  
DB 414 GIFVAVIRINETINKQVDQPNNSIGVLDIFGFENFDNNSFQQLCINANENLQGFVGH 473  
QY 464 IFSLEQLYSRGLVWEDIDWIDGECJDLI-EKKGLALILNESHHPQATDSTLLEKL 522  
DB 474 IFKMEQDEYONHIMQHIEFODNOQIIDLIGMKEMNIMSLIDESKEPKGTDTQTLLEKL 533  
QY 523 HSOHANNHFFYKERYAVANN-FGVKHYAGEVOVDVAGILEKRDPTFDLLNLRSPDF 581  
DB 534 HVQGNRSIYVNGKTTQTSLEIRHYAVGVNPNLGFLEKNDSDSGDRTLVQSTNKKY 593  
QY 582 IYDLFEHVSRRNQDTLKGSGKRRPTVSQFKVDSLMLATSSNPFPRCKIKPMQ 641  
DB 594 LVDIFPH---EMPMOTAK-----KQPLCVAFR-NSLMLWRTLSQAHPIFRCKIPNEY 644  
QY 642 KMPDQFQAVVNLQIRYSGMLETVAIRAGAVARRPFQDFYKRYVLRNLALPE--DVR 699  
DB 645 KEPKFDKELCVRQLRYSGMETARIRAGYPIRHAARYLAVPVRPLQCQCCR 704  
QY 760 GKCTSLQLDYASNSEMOGLTKVFLRESLEQKLEKREBEVSHAMVIRAVLGLFLRK 759  
DB 705 KLAQICVAVLPADSDRGYKTKFLRDEDDASLELRSQMLKSIIVTQIRGVFLFR 764  
QY 760 QYKRVLYCVIILQKRYRAFLLRRLRLHLK-----AALVFQKL----- 798  
DB 765 YMKRYREAIITQRYWRGRLQRRKYQWVRQGNHRLGALIAQQLITKXTMWRCTIYQIA 824  
QY 799 --RGQIARVYRQLLAERKEQEEKKQD-----EEEKKQEEERERER 840  
DB 825 LSRGYLVKRDFOKLLERRKQQLKEELKLAKWKAEBELLRLCQLKEQKEREQEOE 884  
QY 841 EREKRELAQOEERFRKQOELEAOKSKEAELTRELK-----QKEN----- 883  
DB 885 KRLQEEQRLKKAARNALMAAAVQAKRRTYKQEARAPTLQARNSLPPPTTLIVAA 944  
QY 884 ----KQVEIIRL-----EKIEDIQR--MKEQQELS-----LTEAST-- 915  
DB 945 PLPTPRAVAVIRINIPSPSPGIDVSSKQWVDVFRFLNDEPVSGLPGRVYKSKMMP 1004  
QY 916 -QKQERDQELRLREBEACRAAQEFLSLNFEIDECVR-----NIERS---LSGG 963  
DB 1005 EQALRLRDRDVPFKLLSRPAA-----LRKLNNISSGDTIRLPKSVNNIDTDSFYLKVA 1058  
QY 964 SEFSSELAESACEEKNFNFSQPIYEEVDEGFADDDAFQSPRPHSEHSGDQTSGL- 1022  
DB 1059 ATYFGGATQAEHRP-----LKKSLKHEHPIDEMASKAIW 1095  
QY 1023 ----RTSDSSEEDPYNDTV-VPTSPADSTVLLAPVQSGSLHNS--SGESTYCMQ 1076  
DB 1096 LTIARFMD-----LDDVSSPTLHVFNDENLMS-----DLASLNTSDSKRPLFRQ 1144  
QY 1077 NAGDLPSPGDVDYDDDYEDGALITSGS-----VTFNSYG-----SQWSPDY 1120

1145 SCRRIPKPLASGEKEAEFYQHMLNVPSTLSEKIHFIIGHIINKSLRCFPSPSEFEPHL 1204  
QY 1121 RCGVGTYNSSGAVRFSSEG-AOSSFEDSEEDPDSRPTDDELRYRDSVSCUTLIFYPS 1179  
DB 1205 R-----SFMKGTAQLOATPSLQRLERTLVNGPRC--QPPSLBELHAIKGRHP 1250  
QY 1180 FLYMKGALNNSWKRRCVLDKDETFELWFRSKOEBALKO-----GWLH 1219  
DB 1251 -LRDLIDHMDGQGRRLQVDAAST-----AREAVNLCOQMGGLDTFGFGLVMSLNGKLM 1303  
QY 1220 KKGGS-----STLSRRMKRW--FVLROSKLMTFENDSEK-----LK- 1257  
DB 1304 PLGAGEHVLDAISECEORQDAPMKLYRKEMFATWYDPSMDPKATOLYKQILNGLKC 1363  
QY 1258 GTVEVTRAKEIIDNT-----KENGIDIMADRTFHLIASEPD-----ASQWFSVLS 1305  
DB 1364 GEYRCSEKDIAWVACALACVEYGPGEILRLKSEITAFVPSDLLAPGERAIENMSRLIA 1423  
QY 1306 QVHASTDOEIQEMHDEQANPQNAVGTLDGLIDSVCASDSPDRP---NSFVIITANRYL 1361  
DB 1424 ATYEKSSVYEEQNDDLLEAQKRAK-----EDICLFAHLSWPMRHSRLFV----- 1470  
QY 1362 HCNADTPPEMHMTITLQSKSDTRVEGGEFIVRGWLHKE---VKNSPKSSSLKTKR 1416  
DB 1471 --RKEGPKQSDMLMGLINSAGFLIDETEQVLASCCFSEVLKVHVESDDKLVMTFQHV 1528  
QY 1417 WFLYTHNS-----LDYKSSSEKALKGLTLYNSL--CSVPPDEKIFKETGYM 1463  
DB 1529 NFVLQCSSAODANEVINYMIDNLRORSYGVADPVYEGDLEDCLVANPGLIEFAGVT 1588  
QY 1464 NVTYV--GRKCYRLYTKLNEATRWSVYQNTDTKAPIDPTQOOLIODIKENCJLN--- 1518  
DB 1589 GAQLMAGNADCYR---GCYN--GQWQFLAGNVRVLAITKPSKLODILREGRFQBP 1643  
QY 1519 -----SDVVEQIYK-----RNPILRYTHPLASPLPLPYGDINLNL 1556  
DB 1644 KPTPRANYSRRQHNIQSLAESHFREPLDSKAPLSKFSPEPLKAPL----- 1691  
QY 1557 KDKGYTTLODEAIKIFNSLOQ-----LESMSD-----PIP---IIQILQTDHJR 1599  
DB 1692 -----KAVYKVPPPLFOQALVMHHIILKMGDIARSNLPVNTDLIFOPALO--HPL- 1739  
QY 1600 PLDELYCOLIKQTNKVPHPGVSQNLYSQILTCLSTFLPS---RGILKYLKFN--- 1651  
DB 1740 -LCDELYCQMKQLS--DNPSSESEKRGMDLVLATGLVAPSVLVMRELIILMRADAL 1796  
QY 1652 ----LKRIREQPGTEMEKKA-LFTYESLKTKC-----REFVPSRDEIETALHROEMTS 1701  
DB 1797 ADACLRLKRSLSAQGRKAPHLIEVEGIQO--RCLHIYHKIYFPD-DTVEA----- 1845  
QY 1702 TVYCHGGGSCKITINSHTTAGVEVEKLIIGLAMEDSRNMFALFEYNGHVDKAIESRTVA 1761  
DB 1846 -----FEIESHTRGAELIADIQRLK-SPVGYSIFLKTGDRYANPEEFPV 1893  
QY 1762 DVLAK---FEKLAATSEVGDLEPKFYFKLYCFCLDTDNVPKOSV--EFAFMEOAHEAVI 1815  
DB 1894 DFITQIYMLRQORTIRSIDGQYQHFMRKML--NNHGEDLNGDMIFSYQBELHXYL 1951  
QY 1816 HGHHPAPEENLOYLALRYLOGDYTLHAATPLEEYVSLORLKARISOSTKTFPCR 1875  
DB 1952 KGYYPIDCEQASRLAL-----VYSADHVSLORLP----- 1982  
QY 1876 LERKRTSPLEGLRARSFRGVSVRKVEEQMLDMWIKEEVSARASIIDKMRKFGQMNQ 1935  
DB 1983 -----EVLTRILPE---DLIPLOTVAEMRQOILPKVHR-DHLYE 2017  
QY 1936 EQAMATYMLLIKEMPGYSTLPDV-ECKEGFPQELMLGVASADAVSVYKRGGRPLEVFO 1994  
DB 2018 DHAKIIFLOELISFACGSGTFVVKQONDALPETLIINSTGFMHDPPTKEILRSYE 2077  
QY 1995 YEHILSFAGAPLANTYKI---VDERELLFETSEVVDVAKMKAYI 2036  
DB 2078 YSQ-LGIWSSGKNHFHIRGGMIGASKLCTTQGYKMDLLASYV 2122

Search completed: July 14, 2003, 18:15:22  
Job time : 80.3175 secs